**Jurnal Surya Informatika**
**Journal homepage:** https://jurnal.umpp.ac.id/index.php/surya_informatika
P-ISSN: 2477-3042; E-ISSN: 3026-3034
**Vol. 16, No. 1, Mei 2026, Pp. 145-151**

# CIA Triad in the Age of AI: Tinjauan Etika Profesi atas Konflik Anthropic–Pentagon

**Evy Nurmiati*[1)], 'Aisyah Syifa Nur Azzahra[2)]**
1. Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Syarif Hidayatullah Jakarta, Indonesia
2. Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Syarif Hidayatullah Jakarta, Indonesia

**Article Info**
**Kata Kunci:** CIA Triad, etika profesi TI, kecerdasan buatan, keamanan AI, tata kelola AI

***Keywords*:** *AI governance, AI safety, CIA Triad, IT professional ethics, machine learning security*

**Article history:**
Received 21 Mei 2026
Revised 27 Mei 2026
Accepted 28 Mei 2026
Available online 29 Mei 2026

**DOI :**
10.48144/suryainformatika.v16i1.2449

* Corresponding author.
'Aisyah Syifa Nur Azzahra
E-mail address:
aisyah.syifa24@uinjkt.ac.id

**ABSTRAK**
Artikel ini menganalisis transformasi pilar keamanan informasi, *Integrity*, *Confidentiality*, dan *Availability* (*CIA Triad*), dalam konteks sistem kecerdasan buatan (AI) generatif dan *agentic*. *Agentic AI* merujuk pada sistem AI yang mampu mengambil keputusan dan menjalankan tindakan secara semi-otonom tanpa intervensi manusia secara langsung. Dengan menggunakan pendekatan analisis konseptual dan studi kasus kualitatif, penelitian ini mengevaluasi bagaimana ancaman siber kontemporer seperti *prompt injection*, *data poisoning*, dan serangan inferensi merekonfigurasi landasan etika profesi teknologi informasi. Sebagian besar penelitian terdahulu masih membahas *CIA Triad* pada sistem informasi konvensional dan *cloud computing*, sementara kajian dalam konteks AI generatif dan *agentic* serta etika profesi masih terbatas. Studi kasus berfokus pada konflik antara Anthropic dan Departemen Pertahanan Amerika Serikat (Pentagon) pada periode 2025–2026, yang mengekspos ketegangan antara kebijakan keselamatan AI korporasi dan tuntutan operasional militer. Temuan menunjukkan bahwa tanpa kerangka hukum dan tata kelola global yang memadai—sebagaimana disarankan oleh ISO/IEC 42001 dan NIST AI RMF—profesional TI akan terus menghadapi dilema etis yang berpotensi mengancam stabilitas keamanan nasional maupun keselamatan manusia secara luas.

**ABSTRACT**
*This article analyzes the transformation of the information security pillars, Integrity, Confidentiality, and Availability (CIA Triad), within the context of generative and agentic artificial intelligence (AI) systems. Agentic AI refers to systems capable of making decisions and executing actions semi-autonomously toward defined goals without direct human intervention. Using a conceptual analysis and qualitative case study approach, this study examines how contemporary cyber threats such as prompt injection, data poisoning, and inference attacks are reshaping the ethical foundations of the information technology profession. Prior research has predominantly examined the CIA Triad within conventional information systems and cloud computing contexts, while studies addressing generative and agentic AI in relation to professional ethics and defense policy conflicts remain limited. The primary case study focuses on the conflict between Anthropic and the United States Department of Defense (Pentagon) during 2025–2026, which exposed structural tensions between corporate AI safety policies and military operational demands. The findings suggest that without robust legal and global governance frameworks—as recommended by*

**Jurnal Surya Informatika**
**Journal homepage:** https://jurnal.umpp.ac.id/index.php/surya_informatika
P-ISSN: 2477-3042; E-ISSN: 3026-3034
**Vol. 16, No. 1, Mei 2026, Pp. 145-151**
*ISO/IEC 42001 and the NIST AI RMF—IT professionals will continue to face ethical dilemmas that potentially threaten both national security stability and human safety at large.*

## 1. PENDAHULUAN

Perkembangan teknologi informasi telah mencapai titik balik krusial dengan integrasi sistem kecerdasan buatan (AI) generatif dan agentic ke dalam berbagai infrastruktur vital. AI generatif merujuk pada sistem yang mampu menghasilkan konten baru seperti teks, gambar, dan kode secara otomatis, sedangkan agentic AI merujuk pada sistem AI yang memiliki kemampuan mengambil keputusan dan menjalankan tindakan secara semi-otonom berdasarkan tujuan tertentu tanpa intervensi manusia secara langsung. Dalam ekosistem yang semakin terotomatisasi ini, pilar-pilar klasik keamanan informasi yang terangkum dalam CIA Triad, yakni *Integrity*, *Confidentiality*, dan *Availability*, tidak lagi sekadar menjadi parameter teknis, melainkan bertransformasi menjadi fondasi etika profesi yang kompleks. Profesional teknologi informasi (TI) kini memikul tanggung jawab moral yang melampaui pengamanan pangkalan data statis; mereka harus menavigasi risiko di mana algoritma itu sendiri dapat menjadi subjek manipulasi, kebocoran, atau kegagalan operasional yang berdampak luas pada keselamatan publik dan stabilitas nasional. Alief dan Nurmiati (2022) menunjukkan bahwa AI telah berkembang menjadi infrastruktur inti manajemen pengetahuan organisasi, di mana kegagalan atau manipulasi sistem AI berdampak langsung pada kemampuan pengambilan keputusan secara real-time, sehingga memperluas dimensi tanggung jawab etis profesional TI secara signifikan [1]. Sebagian besar penelitian terdahulu masih membahas CIA Triad dalam konteks sistem informasi konvensional, cloud computing, atau keamanan data tradisional. Kajian yang secara eksplisit mengintegrasikan CIA Triad dengan dinamika AI generatif dan agentic, khususnya dalam konteks etika profesi dan konflik kebijakan pertahanan, masih sangat terbatas. Kesenjangan inilah yang menjadi landasan penelitian ini.

Tantangan ini semakin nyata ketika kepentingan korporasi pengembang yang memprioritaskan keselamatan AI berbenturan dengan kebutuhan strategis otoritas pertahanan negara, sebagaimana tercermin dalam konflik antara Anthropic dan Pentagon pada awal tahun 2026 [2]. Konflik tersebut bukan sekadar sengketa kontraktual, melainkan merupakan manifestasi dari ketegangan struktural antara dua sistem nilai yang berbeda: etika keselamatan berbasis korporasi dan logika kesiapan operasional militer.

Evolusi CIA Triad dalam domain AI mencerminkan pergeseran paradigma dari perlindungan data menuju perlindungan proses penalaran mesin. *Integrity* dalam AI modern tidak hanya mencakup keutuhan data pelatihan, tetapi juga keandalan logika model dalam menghasilkan keluaran yang akurat dan bebas dari bias atau manipulasi adversarial. *Confidentiality* kini mencakup perlindungan terhadap parameter model dan pencegahan kebocoran informasi melalui serangan inferensi atau prompt injection. Sementara itu, *Availability* menghadapi ancaman baru berupa penurunan kinerja akibat kebijakan pengamanan yang terlalu restriktif atau serangan siber yang menargetkan infrastruktur komputasi awan [3], [2].

Berdasarkan latar belakang tersebut, artikel ini bertujuan untuk menganalisis transformasi CIA Triad dalam sistem AI generatif dan agenik, mengevaluasi ancaman siber kontemporer yang memengaruhi pilar-pilar CIA, dan membedah dilema etis profesional TI melalui studi kasus konflik Anthropic–Pentagon

## 2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kualitatif dengan dua strategi utama, yaitu analisis konseptual dan studi kasus kualitatif. Pendekatan kualitatif dipilih karena tujuan penelitian bersifat interpretatif dan eksplanatif, yaitu memahami transformasi konsep CIA Triad dalam konteks AI serta menganalisis dinamika etis yang muncul dalam konflik nyata antara aktor teknologi dan pertahanan.

Analisis konseptual dilakukan dengan cara menelaah dan mensintesis literatur akademik, laporan teknis, serta dokumen kebijakan yang relevan dari berbagai sumber terpercaya, meliputi jurnal ilmiah, publikasi lembaga standar internasional (ISO, NIST), laporan industri keamanan siber, serta pemberitaan media yang terverifikasi. Proses ini bertujuan untuk memetakan pergeseran definisi dan implikasi CIA Triad ketika diterapkan pada sistem AI modern.

Studi kasus kualitatif difokuskan pada konflik antara Anthropic dan Departemen Pertahanan Amerika Serikat (Pentagon) pada periode 2025–2026, dengan membandingkan dua perspektif institusional yang berbeda: perspektif korporasi AI (Anthropic) dan perspektif pertahanan militer (Pentagon) terhadap prinsip CIA Triad. Kasus ini dipilih karena merepresentasikan ketegangan yang paling konkret dan terdokumentasi antara prinsip-prinsip keamanan AI korporasi dan tuntutan operasional militer. Data dikumpulkan dari sumber-sumber primer dan sekunder, antara lain pernyataan resmi kedua institusi, analisis hukum dari lembaga hukum internasional (Opinio Juris, Mayer Brown, Pearl Cohen), laporan berita investigatif (CBS News), serta analisis kebijakan dari lembaga think tank (CFR, Cloud Security Alliance) [2], [4], [5], [6], [7], [8].

**Jurnal Surya Informatika**
**Journal homepage:** https://jurnal.umpp.ac.id/index.php/surya_informatika
P-ISSN: 2477-3042; E-ISSN: 3026-3034
**Vol. 16, No. 1, Mei 2026, Pp. 145-151**

Analisis dilakukan secara deskriptif-interpretatif dengan menggunakan kerangka CIA Triad sebagai lensa analitik untuk mengevaluasi setiap dimensi konflik, yaitu *Integrity*, *Confidentiality*, dan *Availability*. Temuan dari analisis konseptual dan studi kasus kemudian disintesis untuk menghasilkan implikasi teoritis dan praktis bagi tata kelola AI dan etika profesi TI.

## 3. HASIL DAN PEMBAHASAN

### *CIA Triad dalam Etika Profesi TI: Landasan Konseptual*

Keamanan sistem informasi modern dibangun di atas model panduan yang dikenal sebagai CIA Triad, yang mencakup tiga komponen utama: *Confidentiality* (Kerahasiaan), *Integrity* (Integritas), dan *Availability* (Ketersediaan). Dalam kajian ini, sistem AI yang dimaksud mencakup AI generatif, yaitu sistem yang mampu menghasilkan konten baru secara otomatis, dan agentic AI, yaitu sistem AI yang mampu mengambil keputusan dan menjalankan tindakan secara semi-otonom berdasarkan tujuan tertentu tanpa intervensi manusia secara langsung [3], [9]. Kerahasiaan didefinisikan sebagai upaya perlindungan informasi dari akses yang tidak sah, memastikan bahwa data sensitif hanya dapat digunakan oleh pihak yang memiliki otoritas [2]. Dalam konteks etika profesi TI, kerahasiaan bukan sekadar kewajiban hukum, tetapi sebuah komitmen moral untuk menghormati privasi pengguna dan melindungi rahasia organisasi[10].

Integritas merujuk pada jaminan bahwa data dan informasi tetap akurat, lengkap, dan tidak diubah oleh pihak yang tidak berwenang selama seluruh siklus hidupnya [3]. Dalam sistem informasi cerdas, integritas meluas ke validitas fungsi algoritmik; apabila logika sebuah model AI dimanipulasi, maka integritas hasil keputusannya dianggap gugur [3]. Ketersediaan memastikan bahwa sistem, aplikasi, dan data dapat diakses oleh pengguna yang berwenang kapanpun dibutuhkan [2]. Aspek ini bersifat kritis dalam skenario di mana ketidaktersediaan layanan dapat menyebabkan kerugian signifikan atau bahkan mengancam nyawa manusia dalam konteks medis maupun militer. Dalam konteks AI sebagai tulang punggung manajemen pengetahuan modern, ketidaktersediaan sistem AI berarti terhentinya seluruh rantai proses pengetahuan organisasi.

Dalam bingkai etika profesi, CIA Triad berfungsi sebagai kompas moral bagi para insinyur dan manajer TI. Kode etik yang diterbitkan oleh *Association for Computing Machinery* (ACM) dan IEEE menegaskan bahwa profesional harus menghindari bahaya dan menjaga kejujuran dalam setiap aspek pekerjaan mereka [10]. Kerangka kerja global seperti ISO/IEC 27001 dan NIST *Cybersecurity Framework* menyediakan metodologi terstruktur untuk mengoperasionalkan prinsip-prinsip ini melalui manajemen risiko yang berkelanjutan. ISO/IEC 42001:2023 melangkah lebih jauh dengan menyediakan standar khusus untuk sistem manajemen AI yang menekankan transparansi, akuntabilitas, dan mitigasi risiko pada teknologi otonom[11].

Tabel 1. Transformasi CIA Triad dari Keamanan Tradisional ke Sistem AI Modern

| Komponen CIA | Keamanan Tradisional | Transformasi dalam Sistem AI Modern |
|---|---|---|
| *Confidentiality* | Perlindungan data mentah dan kredensial akses pengguna. | Perlindungan terhadap model weights, data pelatihan, dan privasi input pengguna dari serangan inferensi. |
| *Integrity* | Pencegahan modifikasi data yang tidak sah melalui hashing dan tanda tangan digital. | Jaminan keandalan output, pencegahan bias algoritma, dan ketahanan terhadap adversarial attacks serta data poisoning. |
| *Availability* | Uptime server, aksesibilitas jaringan, dan pemulihan dari serangan DDoS. | Kecepatan respons model, ketersediaan daya komputasi skala besar, dan keseimbangan antara fungsi safeguards dan performa. |

### *Ancaman Siber Kontemporer terhadap CIA Triad dalam Sistem AI*

Penelitian terbaru mengenai penerapan CIA Triad dalam sistem AI mengungkapkan bahwa ancaman terhadap keamanan informasi telah berevolusi menjadi lebih halus dan manipulatif. Salah satu ancaman paling signifikan adalah prompt injection, yaitu serangan yang memungkinkan penyerang membajak alur logika Large Language Model (LLM) dengan memasukkan instruksi berbahaya yang memaksa model mengabaikan batasan sistemnya[3]. Serangan ini secara langsung mengancam ketiga pilar CIA: membocorkan rahasia sistem (*Confidentiality*), menghasilkan informasi yang salah atau bias (*Integrity*), serta menghentikan layanan melalui eksekusi kode berbahaya (*Availability*)[3].

Kebocoran data dalam AI juga menjadi perhatian utama dalam literatur etika TI. Model AI yang dilatih pada kumpulan data besar seringkali secara tidak sengaja mengingat informasi pribadi yang sensitif. Serangan inferensi dapat digunakan oleh aktor jahat untuk mengekstrak data pelatihan tersebut dari model yang sudah digunakan, yang merupakan pelanggaran berat terhadap prinsip kerahasiaan. Untuk mengatasi ancaman ini, teknik seperti *Privacy-Preserving* AI (PPAI), termasuk *federated learning dan differential privacy*, mulai diadopsi guna meminimalkan transfer data mentah sambil tetap mempertahankan kerahasiaan model.

**Jurnal Surya Informatika**
**Journal homepage:** https://jurnal.umpp.ac.id/index.php/surya_informatika
P-ISSN: 2477-3042; E-ISSN: 3026-3034
**Vol. 16, No. 1, Mei 2026, Pp. 145-151**

Manipulasi model merupakan ancaman serius terhadap integritas yang tidak hanya mencakup modifikasi fisik pada kode, tetapi juga peracunan (poisoning) kumpulan data selama fase pelatihan untuk menyisipkan *backdoor* atau bias tertentu. Jiang et al. (2025) dalam survei sistematisnya terhadap lebih dari 300 studi menyoroti adanya celah generalisasi (*generalization gap*), di mana pertahanan sistem AI sering kali gagal menghadapi ancaman yang terus berevolusi[9]. Hal ini menuntut adanya keamanan intrinsik yang dibangun sejak tahap desain, bukan sekadar sebagai pelengkap pasca-produksi.

Literatur akademik juga menyoroti adanya *trade-off* yang tidak terhindarkan dalam pengembangan AI. Nastoska et al. (2025) menjelaskan bahwa mengoptimalkan satu dimensi kepercayaan seringkali berdampak negatif pada dimensi lainnya [12]. Penerapan mekanisme pengamanan yang sangat ketat dapat menurunkan pengalaman pengguna dan efisiensi operasional. Teknik *differential privacy* sering kali menambahkan noise pada data yang dapat menurunkan akurasi model [12]. Sementara itu, membangun model yang sangat tangguh terhadap serangan adversarial memerlukan komputasi tambahan yang dapat memperlambat waktu respons.

Penyelarasan antara standar keamanan tradisional dan tata kelola AI mulai terbentuk. Greene (2025) menemukan bahwa fungsi-fungsi dalam NIST AI RMF, yakni *Govern, Map, Measure, dan Manage*, sangat selaras dengan fokus CIA Triad dalam kerangka kerja keamanan siber yang telah mapan [11]. Integrasi ini memungkinkan organisasi menggunakan bahasa keamanan yang sama dalam mengelola risiko AI, sehingga integritas data dan ketersediaan sistem tetap menjadi prioritas utama di tengah percepatan inovasi [11].

### *Konflik antara AI Safety, Kebutuhan Operasional, dan Etika Profesional*

Analisis mengenai etika profesi TI dalam pengembangan AI sering bermuara pada konflik antara kebijakan keselamatan AI (*AI safety*) dan kebutuhan operasional pengguna, terutama di sektor pertahanan. Sistem pengamanan (*safeguard*) yang ketat dirancang oleh pengembang untuk mencegah penyalahgunaan AI, seperti penggunaan untuk disinformasi, serangan siber otomatis, atau pengembangan senjata mematikan [2]. Dari perspektif etika profesional, penerapan *safeguard* ini merupakan wujud nyata dari prinsip tanggung jawab terhadap kesejahteraan manusia sebagaimana diamanatkan oleh kode etik ACM [10].

Secara konseptual, *safeguard* yang ketat berfungsi untuk memperkuat pilar *Integrity* dan *Confidentiality* dengan membatasi akses model terhadap fungsi-fungsi berbahaya atau data sensitif, pengembang memastikan bahwa integritas sistem sebagai alat yang bermanfaat tetap terjaga [2]. Namun di sisi lain, batasan ini berpotensi menurunkan *Availability* dari sudut pandang operasional militer. Bagi otoritas pertahanan, AI dipandang sebagai pengganda kekuatan (*force multiplier*) yang harus tersedia tanpa hambatan dari kebijakan internal perusahaan swasta [13]. Ketegangan ini menciptakan dilema di mana keamanan bagi pengembang dimaknai sebagai ketidaktersediaan bagi pengguna.

Pola dilema etis ini tidak hanya terjadi di sektor pertahanan. Khalida et al. (2025) menemukan bahwa lebih dari 74% perguruan tinggi di negara berkembang belum memiliki panduan eksplisit terkait penggunaan AI, mencerminkan kesenjangan tata kelola yang bersifat lintas domain[14]. Hal ini menunjukkan bahwa absennya kerangka etika yang memadai adalah krisis struktural yang melampaui batas sektoral, baik di ranah militer, pendidikan, maupun industri teknologi secara luas.

Konflik ini juga menyentuh aspek loyalitas profesional versus tanggung jawab sosial. Seorang profesional TI memiliki kewajiban kepada pemberi kerja untuk menyediakan sistem yang berfungsi maksimal sesuai kontrak. Namun, kewajiban yang lebih mendasar adalah kepada masyarakat umum untuk tidak menciptakan teknologi yang dapat merugikan kemanusiaan [10]. Tanpa adanya standar etika global yang disepakati, konflik semacam ini akan terus berulang dan mengancam stabilitas tata kelola AI [7], [2].

### *Studi Kasus: Konflik Pentagon–Anthropic dalam Bingkai CIA Triad*

Konflik antara Departemen Pertahanan Amerika Serikat dan Anthropic pada periode 2025–2026 merupakan manifestasi nyata dari ketegangan etis dan keamanan dalam AI modern. Anthropic, yang memposisikan dirinya sebagai perusahaan yang mengutamakan keselamatan, menandatangani kontrak senilai $200 juta dengan Pentagon pada Juli 2025, menjadikan Claude sebagai model AI komersial pertama yang beroperasi di jaringan classified milik Departemen Pertahanan [4]. Dalam kontrak tersebut, Anthropic menegosiasikan dua batasan penggunaan: larangan penggunaan Claude untuk surveilans massa warga negara domestik, dan larangan penggunaan Claude untuk sistem senjata otonom penuh tanpa pengawasan manusia. Perselisihan muncul pada Januari 2026 ketika Pentagon mulai menuntut penghapusan kedua batasan tersebut dan menginginkan akses penuh atas *Claude for all lawful purposes* tanpa pengecualian [2], [5].

### *Integrity: Risiko Manipulasi dan Keandalan Model*

Anthropic berargumen bahwa model AI, termasuk Claude, masih memiliki kerentanan intrinsik terhadap kesalahan keluaran atau halusinasi yang tidak dapat diprediksi [2]. Dalam konteks militer, kegagalan integritas ini dapat berakibat fatal, seperti kesalahan

**Jurnal Surya Informatika**
**Journal homepage:** https://jurnal.umpp.ac.id/index.php/surya_informatika
P-ISSN: 2477-3042; E-ISSN: 3026-3034
**Vol. 16, No. 1, Mei 2026, Pp. 145-151**

identifikasi target atau pelaksanaan serangan yang tidak proporsional [15]. Oleh karena itu, Anthropic bersikeras mempertahankan garis merah yang melarang penggunaan modelnya untuk sistem senjata otonom penuh tanpa pengawasan manusia yang bermakna. Sebaliknya, Pentagon memandang pembatasan ini sebagai penghambat keandalan sistem yang seharusnya bersifat adaptif di medan perang, dengan Kepala Pengadaan Pentagon Emil Michael menyatakan bahwa Anthropic seharusnya "menyeberangi Rubicon" dan membiarkan militer menentukan sendiri batas penggunaan teknologinya [2].

### *Confidentiality: Perlindungan Data dan Pencegahan Surveilans*

Perselisihan mengenai kerahasiaan berpusat pada penolakan Anthropic terhadap penggunaan Claude untuk surveilans massa warga negara Amerika Serikat [2]. Anthropic berargumen bahwa AI mampu mengagregasi data komersial tentang pergerakan, asosiasi, dan perilaku daring warga negara dalam skala dan kecepatan yang tidak dapat diatur oleh hukum privasi yang ada, sehingga menciptakan kapabilitas pengawasan yang tidak akan mungkin terwujud tanpa AI [13]. Pentagon merespons dengan argumen bahwa hukum federal dan kebijakan internal Pentagon sudah melarang surveilans massa dan senjata otonom, sehingga batasan kontraktual Anthropic bersifat redundan dan menempatkan kebijakan perusahaan swasta di atas otoritas negara[2].

### *Availability: Dampak Pembatasan dan Risiko Rantai Pasok*

Eskalasi konflik mencapai puncaknya ketika Menteri Pertahanan Pete Hegseth menetapkan Anthropic sebagai risiko rantai pasokan (*supply chain risk*) pada 27 Februari 2026, sebuah label yang sebelumnya hanya pernah diterapkan pada entitas yang berafiliasi dengan negara asing [5], [4]. Pada hari yang sama, Presiden Trump memerintahkan seluruh lembaga federal untuk segera menghentikan penggunaan produk Anthropic. Penetapan ini berdampak langsung pada ketersediaan layanan Anthropic: mitra pertahanan seperti Palantir dan AWS diwajibkan mensertifikasi bahwa mereka tidak menggunakan Claude dalam pekerjaan yang berkaitan dengan militer, sehingga menciptakan paradoks di mana upaya Anthropic mempertahankan prinsip keselamatan justru menyebabkan layanan mereka tidak tersedia bagi ekosistem pelanggan terbesarnya [6], [2].

### *Dilema Etis dan Implikasi Tata Kelola*

Kasus ini menyisakan dilema etis yang mendalam bagi para profesional pengembang AI. Keputusan CEO Anthropic, Dario Amodei, untuk menolak tuntutan militer meskipun menghadapi risiko kerugian besar, dipuji sebagian kalangan sebagai tindakan integritas profesional [7]. Namun langkah tersebut juga memicu perdebatan bahwa sebuah perusahaan swasta tidak seharusnya memiliki kekuatan untuk menolak keputusan pertahanan nasional yang demokratis [7]. Situasi semakin rumit ketika OpenAI mengumumkan kontrak $200 juta dengan Pentagon hanya beberapa jam setelah Trump menjatuhkan sanksi kepada Anthropic, tanpa memberlakukan batasan serupa, sehingga memunculkan pertanyaan serius tentang konsistensi standar etika di industri AI secara keseluruhan [5].

Tabel 2. Posisi Anthropic dan Pentagon dalam Dimensi Triad

| Komponen CIA | Posisi Anthropic | Posisi Pentagon | Dampak Konflik |
|---|---|---|---|
| *Confidentiality* | Melarang penggunaan AI untuk surveilans massa guna melindungi privasi data warga negara. | Menuntut akses penuh untuk kepentingan intelijen nasional dan pengawasan keamanan domestik. | Potensi gugatan hukum atas pelanggaran privasi dan hak sipil warga negara. |
| *Integrity* | Membatasi AI untuk sistem senjata otonom penuh guna menghindari kesalahan fatal dan ketidakandalan dalam aplikasi kinetik. | Menuntut fleksibilitas penuh dan kemampuan adaptif model di medan perang tanpa pembatasan dari pihak korporasi. | Hilangnya kepercayaan pada reliabilitas model dalam sistem keputusan kritis di lingkungan militer. |
| *Availability* | Menyediakan layanan hanya dalam koridor etika dan keselamatan yang disepakati. | Menetapkan status "*Supply Chain Risk*" untuk memaksakan kepatuhan penuh. | Pemutusan kontrak federal dan pengalihan anggaran ke kompetitor yang lebih kooperatif. |

### *Transformasi Peran Profesional TI dan Masa Depan Tata Kelola AI*

Analisis konflik Anthropic–Pentagon melalui perspektif CIA Triad mengungkap bahwa etika profesi TI di era AI bukan lagi sekadar kepatuhan terhadap aturan teknis, melainkan keterlibatan aktif dalam penentuan batas-batas moral teknologi. Profesional TI kini berada di garis depan dalam menentukan ambang batas risiko yang dapat diterima oleh masyarakat [9]. Kegagalan dalam mempertahankan integritas model dan kerahasiaan data bukan hanya kegagalan teknis, tetapi juga pelanggaran terhadap tanggung jawab sosial yang diamanatkan oleh kode etik internasional [10].

Pengintegrasian AI ke dalam sistem pertahanan nasional menuntut redefinisi terhadap pilar *Availability*. Apabila ketersediaan teknologi harus

**Jurnal Surya Informatika**
**Journal homepage:** https://jurnal.umpp.ac.id/index.php/surya_informatika
P-ISSN: 2477-3042; E-ISSN: 3026-3034
**Vol. 16, No. 1, Mei 2026, Pp. 145-151**

dibayar dengan penghapusan batas-batas etis, maka integritas profesi TI terancam menjadi sekadar instrumen kekuasaan [15]. Studi kasus Anthropic menunjukkan bahwa mekanisme hukum seperti status risiko rantai pasok dapat digunakan sebagai senjata ekonomi untuk memaksa perusahaan teknologi melepaskan prinsip keselamatannya [4].

Tantangan persaingan global, khususnya dari model-model AI Tiongkok yang berkembang pesat seperti Qwen dan DeepSeek, memberikan tekanan tambahan bagi Amerika Serikat untuk tidak membatasi kemampuan AI militernya [7]. Dinamika ini berpotensi menciptakan perlombaan menuju bawah (*race to the bottom*) dalam hal standar keselamatan AI global [10]. Untuk menghadapi tekanan tersebut, profesional TI harus mampu memberikan penilaian yang jujur dan transparan mengenai keterbatasan teknis AI, guna mencegah ketergantungan berlebihan pada sistem otonom yang integritasnya belum teruji sepenuhnya [15].

Jiang et al. (2025) menyoroti bahwa tata kelola AI sering kali dianggap sebagai pemikiran sekunder (afterthought) daripada sebagai prinsip desain yang mendasar [9]. Untuk mengatasi kesenjangan ini, implementasi standar seperti ISO/IEC 42001 perlu menjadi kewajiban bagi penyedia sistem AI di sektor publik, guna memastikan adanya audit etika yang independen dan transparan [11]. Hanya dengan membangun AI yang dapat dipercaya (*trustworthy* AI) yang berlandaskan transparansi dan akuntabilitas, industri teknologi dapat mempertahankan kepercayaan masyarakat di tengah meningkatnya ketegangan geopolitik.

Tabel 3. Metrik Kepercayaan AI dan Relevansinya dalam Kasus Anthropic

| Metric Kepercayaan AI | Deskripsi | Relevansi dalam Kasus Anthropic |
|---|---|---|
| System *Reliability* | Konsistensi operasi sistem sesuai tujuan fungsional yang ditetapkan. | Kekhawatiran Anthropic akan kesalahan fatal AI dalam aplikasi serangan kinetik. |
| Bias Mitigation | Evaluasi keadilan dan ketidakberpihakan hasil keputusan algoritma. | Pencegahan diskriminasi dalam operasi surveilans massa berbasis AI. |
| Model *Explainability* | Transparansi proses pengambilan keputusan model AI yang dapat diaudit. | Kebutuhan audit independen atas tindakan otonom di lingkungan medan perang. |
| *Security Resilience* | Ketahanan sistem terhadap serangan manipulatif dan adversarial. | Pencegahan pembajakan model militer melalui prompt injection dan poisoning. |

## 4. KESIMPULAN

Penerapan prinsip *Integrity*, *Confidentiality*, dan *Availability* dalam sistem kecerdasan buatan modern telah mengalami transformasi yang memaksa redefinisi terhadap etika profesi TI. Integritas tidak lagi sekadar soal keutuhan data, tetapi mencakup keandalan moral dan logis dari sistem otonom. Kerahasiaan meluas dari sekadar privasi menjadi pencegahan penyalahgunaan intelijen berskala besar. Ketersediaan kini menghadapi dilema antara kemudahan operasional dan batasan keselamatan yang diperlukan untuk melindungi kemanusiaan.

Kasus Anthropic versus Pentagon memberikan pelajaran bahwa tanpa kerangka kerja hukum yang kuat untuk melindungi standar keselamatan AI, profesional TI akan terus menghadapi tekanan yang tidak proporsional dari otoritas kekuasaan. Penetapan status risiko rantai pasok terhadap perusahaan yang memegang teguh prinsip etis merupakan peringatan bagi seluruh industri bahwa keselamatan AI adalah isu politik dan hukum yang mendesak, bukan sekadar tantangan teknis. Kesenjangan tata kelola ini, sebagaimana juga terdokumentasi dalam domain pendidikan oleh Khalida et al. (2025) [14], mengonfirmasi bahwa absennya regulasi AI yang komprehensif merupakan tantangan lintas sektor yang mendesak untuk diatasi.

Masa depan tata kelola AI yang etis bergantung pada kemampuan untuk mengintegrasikan prinsip-prinsip CIA Triad ke dalam setiap tahap siklus hidup AI, mulai dari desain hingga pemeliharaan, sebagaimana disarankan oleh standar ISO/IEC 42001 dan NIST AI RMF [11]. Profesional TI harus tetap teguh pada komitmen moral untuk menghindari bahaya, sambil secara aktif berpartisipasi dalam dialog multilateral guna menciptakan standar global yang memastikan kecerdasan buatan tetap menjadi alat yang meningkatkan kesejahteraan manusia.

## 5. UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Syarif Hidayatullah Jakarta atas dukungan akademik yang diberikan selama proses penelitian dan penulisan artikel ini. Penulis juga menyampaikan apresiasi kepada seluruh pihak yang telah memberikan masukan konstruktif, khususnya para reviewer yang telah meluangkan waktu untuk menelaah naskah ini sehingga kualitasnya dapat ditingkatkan. Artikel ini merupakan bagian dari upaya pengembangan kajian etika teknologi informasi di lingkungan perguruan tinggi Islam, dan penulis berharap temuan dalam penelitian ini dapat memberikan kontribusi nyata bagi diskursus akademik seputar tata kelola kecerdasan buatan di Indonesia maupun di tingkat global.

**Jurnal Surya Informatika**
**Journal homepage:** https://jurnal.umpp.ac.id/index.php/surya_informatika
P-ISSN: 2477-3042; E-ISSN: 3026-3034
**Vol. 16, No. 1, Mei 2026, Pp. 145-151**

## REFERENSI

[1] R. Alief and E. Nurmiati, "Penerapan Kecerdasan Buatan Dan Teknologi Informasi Pada Efisiensi Manajemen Pengetahuan," vol. 13, no. 1, pp. 59–69, 2022.

[2] "The Pentagon/Anthropic Clash Over Military AI Guardrails - Opinio Juris." Accessed: May 26, 2026. [Online]. Available: https://opiniojuris.org/2026/02/26/the-pentagon-anthropic-clash-over-military-ai-guardrails/

[3] J. Rehberger, "Trust No AI: Prompt Injection Along The CIA Security Triad," Dec. 2024, Accessed: May 26, 2026. [Online]. Available: https://arxiv.org/pdf/2412.06090

[4] "Pentagon Designates Anthropic a Supply Chain Risk — What Government Contractors Need to Know | Insights | Mayer Brown." Accessed: May 26, 2026. [Online]. Available: https://www.mayerbrown.com/en/insights/publications/2026/03/pentagon-designates-anthropic-a-supply-chain-risk-what-government-contractors-need-to-know

[5] "Pentagon formally designates Anthropic a supply chain risk amid feud over AI guardrails - CBS News." Accessed: May 26, 2026. [Online]. Available: https://www.cbsnews.com/news/pentagon-anthropic-supply-chain-risk-feud-ai-guardrails/

[6] "Anthropic Sues Department of Defense Over Supply Chain Risk Designation - Pearl Cohen." Accessed: May 26, 2026. [Online]. Available: https://www.pearlcohen.com/anthropic-sues-department-of-defense-over-supply-chain-risk-designation/

[7] "Anthropic's Standoff With the Pentagon Is a Test of U.S. Credibility | Council on Foreign Relations." Accessed: May 26, 2026. [Online]. Available: https://www.cfr.org/articles/anthropics-standoff-with-the-pentagon-is-a-test-of-u-s-credibility

[8] "Pentagon vs. Anthropic: Autonomous Weapons AI Guardrails and the Governance Crisis for Enterprise AI Vendors – Lab Space." Accessed: May 26, 2026. [Online]. Available: https://labs.cloudsecurityalliance.org/research/csa-research-note-dod-ai-guardrail-mandates-vendor-governanc/

[9] F. Liu, J. Jiang, Y. Lu, Z. Huang, and J. Jiang, "The ethical security of large language models: A systematic review," *Front. Eng. Manag. 2025 121*, vol. 12, no. 1, pp. 128–140, Jan. 2025, doi: 10.1007/S42524-025-4082-6.

[10] S. Garfinkel, M. Sankaran, R. Sharma, A. B. Shrinivass, A. Pandey, and A. Kumar, "ACM TechBrief: AI-Assisted Software Development, or Vibe Coding: Benefits and Risks of AI-Driven Software Development," Apr. 2026, doi: 10.1145/3807518.

[11] "AI RMF Core - AIRC." Accessed: May 26, 2026. [Online]. Available: https://airc.nist.gov/airmf-resources/airmf/5-sec-core/

[12] N. Díaz-Rodríguez, J. Del Ser, M. Coeckelbergh, M. López de Prado, E. Herrera-Viedma, and F. Herrera, "Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation," *Inf. Fusion*, vol. 99, Nov. 2023, doi: 10.1016/j.inffus.2023.101896.

[13] "Home | Small Wars Journal by Arizona State University." Accessed: May 26, 2026. [Online]. Available: https://smallwarsjournal.com/

[14] R. Khalida, A. Rahmandri, S. Ayla, M. Magren, and E. Nurmiati, "Etika Teknologi Informasi dalam Dunia Pendidikan : Tinjauan Literatur atas Penggunaan AI dan Isu Plagiarisme Akademik melalui Natural Language Processing," vol. 15, no. 2, pp. 222–234, 2025.

[15] J. Johnson, "Can AI behave ethically during military crises? Preserving human moral agency," *Int. Aff.*, vol. 102, no. 1, pp. 63–83, Jan. 2026, doi: 10.1093/IA/IIAF191.