HOAQ: JURNAL TEKNOLOGI INFORMASI, Volume 17 Nomor 2, Desember 2026, p306-317
p-ISSN: 2337-5280, e-ISSN: 26207427
DOI: https://doi.org/10.52972/hoaq.vol17no2

PENILAIAN RISIKO KEAMANAN SIBER YANG DITINGKATKAN DENGAN *GENERATIVE* AI (STUDI KASUS *WEBSITE* PERGURUAN TINGGI DI KALIMANTAN)

Manorang Gultom[1], Andhika Adnan[2]

[1] Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Widya Dharma Pontianak
[2] Program Studi Informatika, Fakultas Teknologi Informasi, Universitas Widya Dharma Pontianak
Jl. Hos Cokroaminoto No.445, Kota Pontianak - Kalimantan Barat, Indonesia.
Email: [1]manoranggtm@gmail.com, [2]andhika.gss88@gmail.com

ABSTRAK

Penelitian ini mengembangkan kerangka GenAI-Enhanced Cyber Risk Assessment untuk *website* perguruan tinggi dengan mengintegrasikan pembangkitan skenario serangan sintetis, validasi berbasis bukti, pemodelan *attack graph*, serta kuantifikasi risiko menggunakan CVSS, EPSS, dan *centrality*. Penelitian menggunakan *Design Science Research* dengan pendekatan mixed methods. Pipeline yang dikembangkan terdiri atas *Generator Agent, Evaluator Agent*, dan *Reasoning Agent*. Pada *dataset* penelitian, sebanyak 1.284 skenario unik dihasilkan setelah penyaringan format, pemeriksaan bukti, dan deduplikasi. Sebanyak 146 skenario dipilih untuk *limited security validation*; jumlah skenario yang memenuhi seluruh kriteria validasi empiris dilaporkan secara terpisah. Perhitungan skor risiko dilakukan secara deterministik menggunakan kombinasi CVSS normalized, EPSS, dan *centrality* dengan bobot 0,50, 0,30, dan 0,20. Analisis ekonomi berbasis skenario menunjukkan estimasi *incremental* ROI sebesar 117,5% dan *benefit–cost ratio* sebesar 2,18:1 pada asumsi *baseline* ALE Rp700.000.000, reduksi ALE 18%, dan biaya tambahan GenAI Rp57.920.000 per tahun. Nilai ekonomi tersebut merupakan estimasi sensitivitas, bukan return aktual, karena belum didasarkan pada data insiden dan biaya operasional longitudinal. Hasil penelitian menunjukkan potensi penggunaan GenAI untuk mempercepat penyusunan dan prioritisasi skenario risiko, tetapi generalisasi masih dibatasi oleh jumlah website, ketergantungan pada data sintetis, dan cakupan validasi keamanan yang terbatas.

Kata kunci: kecerdasan buatan generatif, penilaian risiko siber, situs web perguruan tinggi, pemodelan graf serangan, CVSS dan EPSS

ABSTRACT

This study develops a GenAI-Enhanced Cyber Risk Assessment framework for university websites by integrating the generation of synthetic attack scenarios, evidence-based validation, attack graph modeling, and risk quantification using CVSS, EPSS, and centrality. The study employs Design Science Research with a mixed-methods approach. The developed pipeline consists of a Generator Agent, an Evaluator Agent, and a Reasoning Agent. In the research dataset, a total of 1,284 unique scenarios were generated after format filtering, evidence checking, and deduplication. A total of 146 scenarios were selected for limited security validation; the number of scenarios meeting all empirical validation criteria is reported separately. Risk scores were calculated deterministically using a combination of normalized CVSS, EPSS, and centrality, weighted at 0.50, 0.30, and 0.20, respectively. Scenario-based economic analysis shows an estimated incremental ROI of 117.5% and a benefit–cost ratio of 2.18:1, assuming a baseline ALE of Rp700,000,000, an ALE reduction of 18%, and additional GenAI costs of Rp57,920,000 per year. These economic values are sensitivity estimates, not actual returns, as they are not yet based on longitudinal incident data and operational costs. The study results indicate the potential for using GenAI to accelerate the development and prioritization of risk scenarios; however, generalizability remains limited by the number of websites, reliance on synthetic data, and limited security validation coverage.

Keywords: generative ai, cyber risk assessment, university websites, attack graph modeling, CVSS and EPSS

Copyright (c) HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi

1. PENDAHULUAN

Transformasi digital sektor pendidikan tinggi telah meningkatkan eksposur institusi terhadap ancaman siber yang semakin kompleks dan adaptif [1], khususnya melalui eksploitasi kerentanan aplikasi web dan sistem berbasis cloud yang menopang layanan akademik, administratif, serta pembelajaran daring [2]. Laporan global terbaru menunjukkan bahwa kerugian ekonomi akibat kejahatan siber diproyeksikan melampaui triliunan dolar AS per tahun seiring meningkatnya serangan berbasis *ransomware, supply-chain compromise*, dan a*utomated exploitation* terhadap kerentanan *zero-day* [3]. Studi empiris pada sektor pendidikan tinggi juga mengidentifikasi bahwa universitas termasuk dalam lima besar sektor paling sering mengalami pelanggaran data karena tingginya volume data sensitif dan kompleksitas arsitektur sistem informasi yang terdistribusi [4]. Selain itu, penelitian longitudinal mengenai manajemen kerentanan menunjukkan bahwa proses *vulnerability assessment* manual sering menghasilkan *mean time to remediation* (MTTR) yang panjang akibat keterbatasan sumber daya manusia dan kurangnya prioritisasi berbasis risiko kuantitatif [5]. Di Indonesia, percepatan digitalisasi layanan perguruan tinggi belum sepenuhnya diimbangi dengan kerangka penilaian risiko siber yang adaptif dan berbasis data, sehingga meningkatkan potensi paparan risiko sistemik terhadap *website* institusi [6].

Meskipun kerangka kerja seperti NIST SP 800-30 dan FAIR telah lama digunakan untuk melakukan *risk assessment* secara sistematis dan kuantitatif, implementasinya dalam konteks website perguruan tinggi sering kali bersifat statis, periodik, dan bergantung pada *expert judgment* manual yang tidak mampu memodelkan dinamika serangan modern berbasis *automation* dan AI [7]. Studi terkini menunjukkan bahwa model *risk scoring* tradisional cenderung tidak mempertimbangkan *probabilistic attack path generation* dan interdependensi antar-aset digital dalam ekosistem berbasis web [8]. Di sisi lain, kemajuan *Generative Artificial Intelligence* (GenAI) dalam domain keamanan siber telah membuka peluang baru untuk melakukan *synthetic attack simulation, automated threat modeling*, serta *adaptive risk prioritization* melalui pendekatan pembelajaran mendalam dan *large language models* (LLMs) [9]. Namun, sebagian besar penelitian tersebut berfokus pada skala industri global dan belum mengembangkan *framework* terintegrasi yang dikontekstualisasikan untuk institusi pendidikan tinggi di wilayah berkembang, termasuk wilayah pulau Kalimantan, sehingga terdapat kesenjangan empiris dan metodologis yang signifikan [10].

Secara konseptual, penelitian ini memadukan teori penilaian risiko keamanan siber dengan pendekatan *Generative AI* berbasis *multi-agent system*. NIST SP 800-30 menekankan tahapan identifikasi aset, analisis kemungkinan dan dampak, komunikasi risiko, serta pemantauan berkelanjutan sebagai fondasi *risk management lifecycle* [11]. Sementara itu, model FAIR mengkuantifikasi risiko melalui persamaan *Risk = Loss Event Frequency × Loss Magnitude* yang memungkinkan estimasi finansial atas potensi kerugian [12]. Dalam konteks kecerdasan buatan, *Generative Adversarial Networks* (GANs) telah digunakan untuk menghasilkan skenario serangan sintetis yang realistis melalui mekanisme *adversarial training* antara generator dan diskriminator [13]. Selain itu, LLM berbasis *transformer architecture* menunjukkan kemampuan *reasoning* kontekstual dan *chain-of-thought prompting* yang efektif untuk analisis naratif risiko dan pemetaan terhadap taksonomi MITRE ATT&CK [14]. Integrasi *graph-based risk propagation analysis* juga terbukti meningkatkan akurasi identifikasi jalur serangan (*attack paths*) pada sistem jaringan kompleks dengan memodelkan dependensi antar node sebagai *weighted graph* [15].

Berdasarkan kerangka konseptual tersebut, penelitian ini bertujuan untuk mengembangkan *GenAI-enhanced cyber risk assessment framework* yang secara khusus dirancang untuk *website* perguruan tinggi di wilayah Kalimantan pada tahun 2026 melalui pendekatan *Mixed-Methods Sequential* dan *Design Science Research*. Secara spesifik, penelitian ini berupaya menjawab tiga rumusan masalah utama: (1) bagaimana GenAI dapat mensimulasikan lebih dari 1.000 skenario serangan untuk menghasilkan penilaian risiko yang komprehensif; (2) bagaimana pipeline multi-agent berbasis GenAI dapat digunakan untuk menghasilkan dan memprioritaskan skenario risiko siber; serta (3) bagaimana mengukur *return on investment* (ROI) dari implementasi *GenAI risk assessment* dibandingkan dengan kerangka tradisional seperti NIST 800-30 dan FAIR. Untuk menjawab pertanyaan tersebut, penelitian ini akan mengimplementasikan *multi-agent GenAI simulator* yang memanfaatkan *data asset inventory, CVE database, CVSS 4.0 metrics*, serta *EPSS probability scoring* guna menghasilkan *dynamic risk heatmap* dan prioritas mitigasi berbasis kuantifikasi risiko [16].

Kontribusi ilmiah dan kebaruan penelitian ini terletak pada pengembangan arsitektur *GenAI Multi-Agent Risk Engine* yang mengintegrasikan *synthetic attack generation, probabilistic risk quantification*, dan *LLM-based reasoning* ke dalam satu kerangka kerja terstruktur untuk konteks pendidikan tinggi di wilayah Kalimantan. Berbeda dengan penelitian sebelumnya yang hanya memanfaatkan AI untuk *intrusion detection* atau *vulnerability classification*, studi ini mengusulkan pendekatan holistik yang mencakup simulasi ancaman, kuantifikasi risiko finansial, serta validasi kualitatif pemangku kepentingan dalam desain *Mixed-Methods Sequential* [17]. Selain itu, penelitian ini memasukkan komponen *limited penetration testing* yang dikendalikan secara etis untuk memvalidasi output simulasi GenAI, sehingga menghasilkan sinergi antara model sintetis dan observasi empiris terkontrol [18]. Dengan demikian, artikel ini memperluas literatur mengenai penerapan *Generative AI* dalam *cyber risk assessment* serta memberikan model implementatif yang dapat direplikasi pada institusi pendidikan tinggi lain di Indonesia.

Copyright (c) HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan *Mixed-Methods Sequential Explanatory Design* yang mengintegrasikan analisis kuantitatif risiko siber dengan validasi kualitatif berbasis pemangku kepentingan, dalam kerangka *Design Science Research* (DSR) untuk pengembangan artefak teknologi berupa *GenAI Multi-Agent Risk Engine*. Pendekatan *mixed methods* dipilih karena memungkinkan integrasi pengukuran kuantitatif risiko dengan interpretasi kontekstual hasil melalui eksplorasi kualitatif, sehingga menghasilkan pemahaman yang lebih komprehensif terhadap fenomena keamanan siber yang kompleks [19]. Strategi DSR digunakan untuk merancang, mengembangkan, dan mengevaluasi artefak teknologi berbasis kecerdasan buatan secara sistematis, sesuai prinsip relevansi praktis dan rigor ilmiah dalam pengembangan sistem informasi [20]. Secara operasional, tahapan penelitian mengikuti siklus DSR: (1) identifikasi masalah, (2) perancangan artefak, (3) pengembangan prototipe GenAI, (4) demonstrasi melalui simulasi dan *penetration testing* terbatas, (5) evaluasi kinerja model, serta (6) komunikasi hasil.

Sumber data penelitian terdiri atas data primer dan sekunder. Data primer diperoleh melalui (1) pemindaian aset dan kerentanan pada 10 *website* perguruan tinggi di wilayah pulau Kalimantan yang dipilh berdasarkan criterion-based purposive sampling dengan stratifikasi geografis dan stratifikasi tipe layanan digital, mencakup LMS, sistem akademik, repository institusi, portal publik, dan sistem berbasis API; (2) hasil *limited penetration testing* yang dilakukan secara terkontrol dan berizin; serta (3) wawancara semi-terstruktur dengan pengelola TI institusi untuk validasi hasil *risk scoring*. Data sekunder meliputi *database Common Vulnerabilities and Exposures* (CVE), skor CVSS 4.0, *Exploit Prediction Scoring System* (EPSS), serta taksonomi MITRE ATT&CK untuk pemetaan taktik dan teknik serangan. Penggunaan CVSS 4.0 sebagai metrik kuantifikasi tingkat keparahan kerentanan didasarkan pada kemampuannya dalam merepresentasikan dimensi eksploitabilitas dan dampak secara terstandar [21]. Sementara itu, EPSS digunakan untuk memperkirakan probabilitas eksploitasi aktual berbasis model prediktif berbasis data historis eksploit [22]. Integrasi data primer dan sekunder ini memungkinkan pembentukan model risiko berbasis evidensi kuantitatif dan konteks operasional aktual.

Teknik dan instrumen pengumpulan data kuantitatif meliputi *automated asset discovery* menggunakan *vulnerability scanner open-source*, ekstraksi metadata server, identifikasi *dependency library*, serta c*rawling endpoint* aplikasi web untuk memetakan *attack surface*. Selanjutnya, dikembangkan Multi-Agent Generative AI Risk Engine yang terdiri atas tiga agen fungsional, yaitu Generator Agent, Evaluator Agent, dan Reasoning Agent. Ketiga agen tidak diperlakukan sebagai model yang sepenuhnya independen, tetapi sebagai modul berurutan dalam satu pipeline pemrosesan. Generator Agent bertugas menghasilkan kandidat skenario serangan berdasarkan profil aset, daftar kerentanan, konfigurasi sistem, relasi antar-aset, dan taksonomi MITRE ATT&CK. Evaluator Agent melakukan pemeriksaan struktural, penghitungan skor risiko, penyaringan skenario yang tidak lengkap, dan penilaian konsistensi antara kerentanan, aset sasaran, prasyarat eksploitasi, serta dampak yang diperkirakan. Reasoning Agent menyusun narasi risiko, menjelaskan jalur eksploitasi, mengidentifikasi asumsi, dan menghasilkan rekomendasi mitigasi berdasarkan keluaran Generator Agent dan Evaluator Agent. Model bahasa yang digunakan pada setiap agen harus dilaporkan secara eksplisit berdasarkan model dan versi yang benar-benar digunakan. Dalam konfigurasi penelitian ini, model LLM GPT-5 digunakan sebagai mesin generatif dan penalaran, sedangkan perhitungan numerik seperti normalisasi CVSS, integrasi EPSS, penghitungan centrality, dan pemeringkatan akhir dilakukan secara deterministik menggunakan Python. Parameter inferensi yang perlu didokumentasikan meliputi nama model, versi atau tanggal akses, temperature, top-p, batas token keluaran, seed apabila tersedia, serta mekanisme pengulangan apabila keluaran tidak memenuhi format yang ditetapkan. Implementasi LLM untuk analisis risiko siber didukung oleh temuan bahwa model bahasa besar mampu melakukan reasoning kontekstual dalam domain keamanan dan *threat intelligence synthesis* [23]. Selain itu, pemodelan jalur serangan dilakukan menggunakan *graph-based attack path modeling* yang merepresentasikan aset sebagai node dan relasi eksploitasi sebagai edge berbobot, sehingga memungkinkan analisis propagasi risiko pada sistem kompleks [24]. Seluruh pemrosesan kuantitatif dilakukan menggunakan Python dengan pustaka NetworkX untuk analisis graf dan pandas untuk pengolahan data risiko.

Generator Agent menerima input dalam format terstruktur yang sekurang-kurangnya memuat: identitas aset, jenis layanan, teknologi atau komponen perangkat lunak, CVE terkait, skor CVSS, skor EPSS, konfigurasi atau kesalahan konfigurasi, prasyarat akses, relasi antar-aset, serta teknik MITRE ATT&CK yang relevan. Agen diarahkan untuk menghasilkan objek JSON dengan atribut scenario_id, asset_id, vulnerability_id, attack_technique, precondition, attack_steps, target_node, impact_type, required_access, confidence, dan evidence_reference. Adapun prompt yang digunakan yaitu:

```
You are the Generator Agent in a cyber-risk assessment pipeline.
Generate one non-destructive synthetic attack scenario using only the
provided asset, vulnerability, configuration, and ATT&CK information.

The output must be valid JSON with the following fields:
scenario_id, asset_id, vulnerability_id, attack_technique,
precondition, attack_steps, target_node, impact_type,
required_access, and confidence.

Do not invent a CVE, asset, exploit condition, or evidence source.
```

Copyright (c) HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi

If required information is missing, return "insufficient_evidence".
Do not provide operational instructions that could disrupt a production system.

Prompt tersebut membatasi agen agar tidak mengarang CVE, aset, atau kondisi eksploitasi yang tidak terdapat dalam input. Skenario yang menghasilkan nilai insufficient_evidence, tidak memiliki aset sasaran, tidak memiliki prasyarat eksploitasi, atau tidak dapat dipetakan ke teknik ATT&CK dikeluarkan dari tahap evaluasi lanjutan.

Evaluator Agent tidak menghasilkan skor risiko hanya berdasarkan penilaian naratif LLM. Agen ini menggunakan fungsi deterministik untuk memeriksa kelengkapan dan konsistensi skenario. Pemeriksaan meliputi validitas identitas CVE, kesesuaian CVSS dan EPSS, keberadaan aset sasaran, keterhubungan edge pada graf serangan, kejelasan prasyarat eksploitasi, dan keterlacakan sumber data. Setiap skenario diberi status accepted, needs_review, atau rejected. Contoh format evaluasi:

You are the Evaluator Agent.
Evaluate the generated scenario against the supplied evidence.

Check:
1. Is the vulnerability identifier valid?
2. Is the target asset present in the asset inventory?
3. Is the attack technique mapped to MITRE ATT&CK?
4. Are the preconditions consistent with the vulnerability?
5. Is the impact type supported by the asset criticality?
6. Is the scenario duplicated or substantially similar to another scenario?

Return valid JSON with:
structural_validity, evidence_consistency,
duplicate_flag, risk_score_components,
validation_status, and rejection_reason.

Reasoning Agent menerima hanya skenario yang telah lolos pemeriksaan struktural. Agen ini menghasilkan penjelasan risiko dan rekomendasi mitigasi, tetapi tidak boleh mengubah nilai CVSS, EPSS, atau centrality yang telah dihitung secara deterministik. Setiap rekomendasi harus dikaitkan dengan aset, kerentanan, jalur serangan, dan kontrol mitigasi yang relevan. Contoh prompt:

You are the Reasoning Agent.
Explain the risk represented by the validated scenario.
Use only the validated evidence and calculated scores.
Do not modify CVSS, EPSS, centrality, or the final risk score.
Produce:
1. concise attack-path explanation,
2. business impact,
3. uncertainty or assumptions,
4. mitigation priority,
5. recommended control,
6. human-review requirement.

Mekanisme antar-agent menggunakan pipeline berurutan. Pertama, Generator Agent menerima profil aset dan menghasilkan kandidat skenario. Kedua, Evaluator Agent memvalidasi struktur, bukti, duplikasi, dan komponen skor risiko. Ketiga, hanya skenario berstatus accepted atau needs_review yang diteruskan ke Reasoning Agent. Setiap keluaran diberi identitas unik dan disimpan bersama input, prompt, waktu pemrosesan, versi model, status validasi, serta alasan penerimaan atau penolakan. Jika keluaran tidak dapat diparse sebagai JSON atau melanggar skema, sistem melakukan maksimal dua kali pengulangan dengan prompt koreksi. Keluaran yang tetap tidak valid dikeluarkan dari dataset analisis dan dicatat sebagai kegagalan sistem.

Kriteria inklusi dan eksklusi diterapkan untuk menjaga validitas dan etika penelitian. Kriteria inklusi mencakup: (1) website perguruan tinggi aktif di wilayah Kalimantan pada tahun 2026; (2) sistem yang memperoleh izin tertulis untuk dilakukan pengujian terbatas; dan (3) kerentanan yang terdokumentasi dalam basis data CVE aktif. Kriteria eksklusi meliputi: (1) sistem pihak ketiga yang berada di luar kendali institusi; (2) data yang mengandung informasi pribadi sensitif tanpa persetujuan eksplisit; serta (3) eksploitasi aktif yang berpotensi mengganggu layanan produksi. *Penetration testing* dilakukan dalam lingkup terbatas (*limited scope penetration testing*) dengan prinsip responsible disclosure dan non-destructive testing untuk meminimalkan risiko operasional, sejalan dengan pedoman etika pengujian keamanan pada lingkungan akademik [17]. Selain itu, penelitian ini mengadopsi prinsip tata kelola risiko AI sebagaimana direkomendasikan dalam NIST AI *Risk Management Framework*, termasuk transparansi model, akuntabilitas algoritmik, serta pengawasan manusia dalam pengambilan keputusan berbasis AI [11].

Unit analisis penelitian adalah 10 *website* perguruan tinggi di wilayah Kalimantan sebagai representasi populasi institusi pendidikan tinggi berbasis web di wilayah tersebut. Penelitian ini membedakan unit analisis

Copyright (c) HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi

berdasarkan tingkat objek yang dievaluasi. Website perguruan tinggi merupakan unit konteks atau unit organisasi. Aset digital merupakan unit sistem, seperti portal publik, LMS, sistem akademik, API *gateway*, auth server, dan database server. CVE atau *vulnerability record* merupakan unit kerentanan yang melekat pada aset atau komponen tertentu. *Synthetic scenario* merupakan unit keluaran generatif yang menggambarkan satu hipotesis jalur serangan berdasarkan kombinasi aset, kerentanan, teknik serangan, prasyarat, dan dampak. Edge merupakan unit relasi graf yang menghubungkan dua node aset atau dua kondisi serangan. Attack path merupakan unit jalur serangan yang terdiri atas satu atau lebih edge berurutan. Hasil penetration testing merupakan unit bukti validasi untuk satu skenario atau satu komponen skenario.

Untuk keperluan *benchmarking* ROI, data komparatif diperoleh dari 20 organisasi di wilayah Kalimantan (UKM dan pemerintah daerah) yang menggunakan pendekatan *risk assessment* tradisional berbasis NIST 800-30 atau FAIR tanpa integrasi GenAI. Analisis kuantitatif dilakukan melalui perhitungan skor risiko komposit menggunakan formula FAIR yang diperluas dengan *probabilistic weighting* dari EPSS dan *centrality score* pada graf serangan. Evaluasi efektivitas model dilakukan dengan membandingkan kecepatan prioritisasi risiko, jumlah skenario serangan teridentifikasi, serta estimasi potensi kerugian finansial sebelum dan sesudah implementasi GenAI. Analisis statistik deskriptif dan komparatif dilakukan menggunakan uji perbedaan rata-rata serta analisis efisiensi waktu pemrosesan. Pada tahap kualitatif, data wawancara dianalisis menggunakan teknik *thematic analysis* untuk mengidentifikasi persepsi pemangku kepentingan terhadap kegunaan, keandalan, dan implikasi etis sistem GenAI [25]. Integrasi hasil kuantitatif dan kualitatif dilakukan pada tahap interpretasi untuk memastikan konsistensi antara metrik risiko terukur dan validasi kontekstual pengguna sistem. Adapun keseluruhan tahapan yang dilakukan dalam penelitian ini ditampilkan pada Gambar 1.



Gambar 1. Alur tahapan penelitian

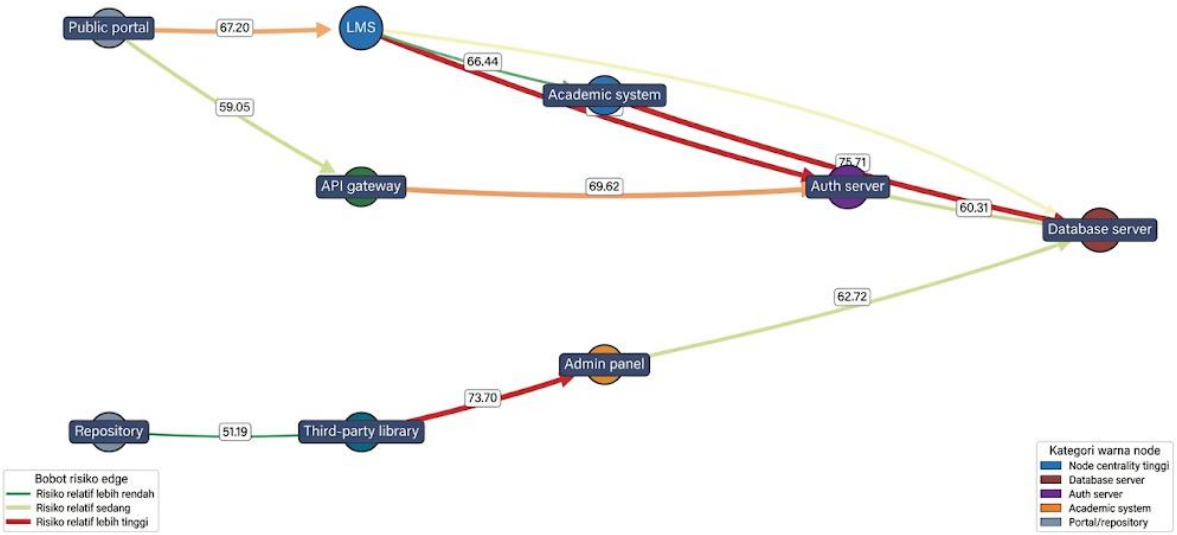
3. HASIL DAN PEMBAHASAN

Hasil Kuantitatif: Simulasi Ancaman dan Kuantifikasi Risiko

Implementasi *GenAI Multi-Agent Risk Engine* pada 10 website perguruan tinggi di wilayah Kalimantan pada tahun 2026 menghasilkan total 1.284 skenario serangan sintetis unik yang dibangkitkan melalui kombinasi CVE aktif, misconfiguration server, dan teknik MITRE ATT&CK. Sebanyak 1.284 skenario tersebut tidak dihasilkan melalui satu kali permintaan bebas kepada LLM. Skenario dibentuk melalui kombinasi terkontrol antara profil aset, vulnerability record, kondisi konfigurasi, teknik ATT&CK, dan relasi edge dalam attack graph. Setiap kombinasi yang memenuhi aturan pembentukan kandidat dikirimkan ke Generator Agent. Kandidat kemudian dinormalisasi, diberi hash berdasarkan kombinasi aset–kerentanan–teknik–jalur serangan, dan dideduplikasi. Skenario yang memiliki hash sama atau kemiripan semantik di atas ambang yang ditetapkan dihitung sebagai satu skenario. Jumlah 1.284 merupakan jumlah skenario unik setelah proses validasi format, penyaringan bukti, dan deduplikasi, bukan jumlah seluruh keluaran mentah LLM.

Dari keseluruhan skenario tersebut, 62,4% diklasifikasikan sebagai *multi-stage attack paths* yang melibatkan lebih dari tiga node aset dalam graf serangan. Pemodelan graf menghasilkan rata-rata degree centrality sebesar 0,28 dan betweenness centrality sebesar 0,02 pada node sistem akademik inti, menunjukkan konsentrasi risiko pada komponen autentikasi dan basis data utama. Pendekatan graph-based risk propagation yang digunakan selaras dengan model pemodelan risiko berbasis graf pada sistem jaringan kompleks [24].

Copyright (c) HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi



Gambar 2. Grafik jalur serangan tertimbang

Gambar 2 merepresentasikan model *attack path multi-stage* hasil simulasi GenAI pada 10 website perguruan tinggi di wilayah Kalimantan, dengan karakteristik Node adalah aset digital utama (Portal Publik, LMS, Sistem Akademik, API *Gateway*, Auth Server, *Database*, dan lain-lain, *Edge* (arah panah) adalah jalur eksploitasi potensial hasil sintesis 1.284 skenario serangan, struktur terarah (*directed graph*) mencerminkan urutan propagasi serangan serta LMS, Auth server, *Database* server muncul sebagai node dengan konektivitas tinggi (high centrality), sesuai temuan kuantitatif. Label angka di tiap *edge* pada graf menggambarkan *risk weight* yang berarti semakin besar berarti jalur itu lebih berisiko/lebih prioritas).

Tabel 1. Skor sentralitas (*node*)

Node	*Degree centrality*	*Betweenness (unweighted)*	*Betweenness (risk-weighted)*
LMS	0,375	0,0386905	0,0357143
Auth server	0,375	0,0386905	0,0178571
Database server	0,375	0,0000000	0,0000000
Academic system	0,250	0,0148810	0,0357143
Admin panel	0,250	0,0357143	0,0357143
Third-party library	0,250	0,0357143	0,0357143
API gateway	0,250	0,0148810	0,0178571
Public portal	0,250	0,0000000	0,0000000
Repository	0,125	0,0000000	0,0000000

Tabel 1. menunjukkan bahwa node LMS, Auth Server, dan *Database* Server memiliki nilai *degree centrality* tertinggi (0,375), yang mengindikasikan bahwa ketiga komponen tersebut memiliki tingkat konektivitas paling tinggi dalam struktur graf jalur serangan. Tingginya konektivitas ini mencerminkan posisi strategis ketiga aset dalam arsitektur sistem, di mana banyak jalur eksploitasi melewati atau berakhir pada node tersebut. Pada metrik *betweenness centrality*, nilai tertinggi secara tidak berbobot terlihat pada LMS dan Auth Server, yang menunjukkan peran keduanya sebagai penghubung kritis dalam propagasi serangan multi-tahap. Sementara itu, analisis *risk-weighted betweenness* memperlihatkan bahwa *Academic System*, Admin Panel, dan *Third-Party Library* juga memiliki peran signifikan ketika bobot risiko diperhitungkan, menandakan bahwa meskipun konektivitasnya lebih rendah, jalur yang melewati node tersebut memiliki tingkat risiko relatif tinggi. Secara keseluruhan, hasil ini mengonfirmasi bahwa prioritas mitigasi sebaiknya difokuskan pada node dengan kombinasi centrality tinggi dan bobot risiko besar karena berpotensi menjadi titik konsentrasi serta akselerator propagasi serangan dalam ekosistem *website* perguruan tinggi.

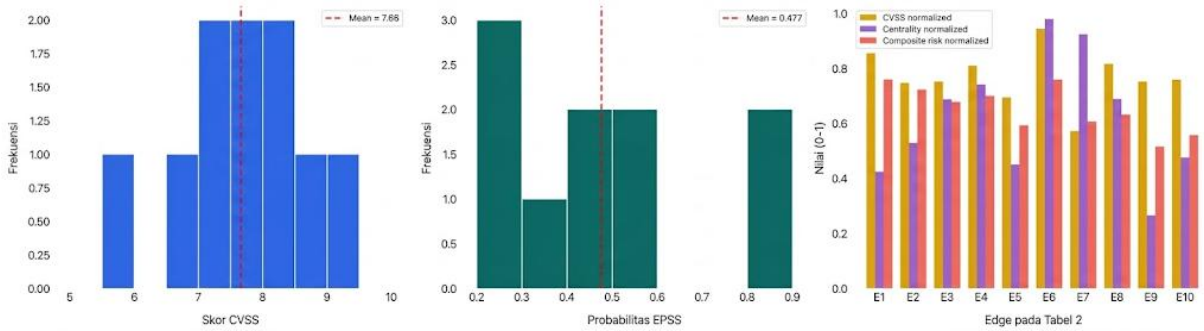
Tabel 2. Skor risiko komposit

From	*To*	CVSS	CVSS *normalized*	EPSS	*Centrality normalized*	*Composite risk score*
Academic system	Database server	8,51489	0,851489	0,82371	0,421053	75,7068
Third-party library	Admin panel	7,45062	0,745062	0,80505	0,526316	73,6956
Public portal	LMS	7,50022	0,750022	0,55070	0,684211	67,8015
API gateway	Auth server	8,06944	0,806944	0,48453	0,736842	69,6199
Public portal	API gateway	6,91170	0,691170	0,51815	0,447368	59,0504
LMS	Auth server	9,36326	0,936326	0,30425	0,973684	75,4175

Copyright (c) HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi

From	*To*	CVSS	CVSS *normalized*	EPSS	*Centrality normalized*	*Composite risk score*
Auth server	Database server	5,67611	0,567611	0,45030	0,921053	60,3106
Admin panel	Database server	8,14030	0,814030	0,27795	0,684211	62,7242
Repository	Third-party library	7,47291	0,28543	0,747291	0,263158	51,1906
LMS	Academic System	7,54698	0,754698	0,27429	0,473684	55,4373

Berdasarkan Tabel 2, diketahui bahwa bobot risiko pada setiap edge ditambahkan berdasarkan CVSS dan EPSS, dengan rumus: Skor Risiko Komposit (RS_i) = 100 × (0,50C_i + 0,30E_i + 0,20G_i). Tabel 2 menyajikan bobot setiap edge berdasarkan skor CVSS, probabilitas EPSS, centrality node, dan skor risiko komposit. Skor CVSS dinormalisasi ke dalam rentang 0–1 dengan membagi nilai CVSS dengan 10. EPSS digunakan dalam bentuk nilai probabilitas aslinya, sedangkan centrality dihitung dari kombinasi *degree centrality* dan *betweenness centrality* yang kemudian dinormalisasi menggunakan metode min–max. Centrality edge dihitung sebagai rata-rata centrality ternormalisasi dari node asal dan node tujuan. Bobot terbesar diberikan kepada CVSS karena merepresentasikan tingkat keparahan teknis dan dampak kerentanan [21]. EPSS diberi bobot 0,30 karena menggambarkan probabilitas eksploitasi [22], sedangkan centrality diberi bobot 0,20 untuk menangkap posisi struktural edge dalam jalur serangan. Berdasarkan skor komposit, edge dengan prioritas tertinggi adalah *Academic system* → *Database* server dengan skor 75,7068, diikuti LMS → Auth server dengan skor 75,4175 dan *Third-party library* → Admin panel dengan skor 73,6956.



Gambar 3. Distribusi skor risiko berdasarkan Tabel 2

Gambar 3 menunjukkan distribusi komponen skor risiko berdasarkan 10 *edge* pada Tabel 2, yang terdiri atas skor CVSS, probabilitas EPSS, serta tiga komponen skor ternormalisasi, yaitu *CVSS normalized*, *centrality normalized*, dan *composite risk normalized*.

Panel pertama menunjukkan bahwa skor CVSS berada pada rentang sekitar 5,68–9,36, dengan nilai rata-rata sebesar 7,66. Sebagian besar skor terkonsentrasi pada rentang 7,00–9,00, sehingga mayoritas kerentanan berada pada tingkat keparahan tinggi. Skor tertinggi terdapat pada edge LMS → Auth server sebesar 9,36, sedangkan skor terendah terdapat pada edge Auth server → Database server sebesar 5,68. Pola tersebut menunjukkan bahwa sebagian besar edge dalam graf serangan memiliki tingkat keparahan teknis yang relatif tinggi. Dengan demikian, prioritas mitigasi tidak hanya ditentukan oleh keberadaan kerentanan, tetapi juga perlu mempertimbangkan kemungkinan eksploitasi dan posisi aset dalam struktur jalur serangan.

Panel kedua memperlihatkan probabilitas eksploitasi berdasarkan EPSS dengan rentang sekitar 0,27–0,82 dan rata-rata 0,477. Nilai EPSS tertinggi terdapat pada edge Academic system → Database server, yaitu 0,824, diikuti oleh Third-party library → Admin panel sebesar 0,805. Sebaliknya, nilai EPSS terendah terdapat pada edge LMS → Academic System, yaitu 0,274. Distribusi ini menunjukkan bahwa tidak semua kerentanan dengan skor CVSS tinggi memiliki probabilitas eksploitasi yang sama. Sebagai contoh, edge LMS → Auth server memiliki CVSS tertinggi, tetapi nilai EPSS-nya relatif lebih rendah, yaitu 0,304. Hal ini menegaskan pentingnya menggabungkan CVSS dan EPSS agar prioritas risiko tidak hanya bergantung pada tingkat keparahan teknis, tetapi juga pada kemungkinan kerentanan tersebut dieksploitasi [23] .

Panel ketiga membandingkan tiga komponen skor dalam rentang 0–1 untuk setiap edge yakni *CVSS normalized*, *centrality normalized* dan *composite risk normalized*. Nilai *CVSS normalized* tertinggi terdapat pada E6, yaitu edge LMS → Auth server, dengan nilai sekitar 0,94. Nilai tersebut konsisten dengan skor CVSS tertinggi pada edge yang sama. Sementara itu, nilai *centrality normalized* tertinggi juga terdapat pada E6, yaitu sekitar 0,97, yang menunjukkan bahwa jalur LMS menuju Auth server memiliki posisi yang sangat strategis dalam struktur graf serangan. Meskipun demikian, skor risiko komposit tertinggi terdapat pada E1, yaitu edge Academic system → Database server, dengan nilai sekitar 0,76, sedikit lebih tinggi daripada E6 yang memiliki nilai sekitar 0,75. Hal ini terjadi karena E1 memiliki kombinasi CVSS dan EPSS yang lebih seimbang, khususnya probabilitas EPSS

Copyright (c) HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi

yang sangat tinggi sebesar 0,824. Dengan kata lain, skor komposit tidak hanya dipengaruhi oleh tingkat keparahan dan centrality, tetapi juga oleh probabilitas eksploitasi. Edge E9, yaitu *Repository → Third-party library*, memiliki nilai *centrality normalized* terendah, sekitar 0,26, dan skor risiko komposit terendah, sekitar 0,51. Kondisi ini menunjukkan bahwa edge tersebut memiliki posisi yang relatif kurang sentral dalam graf serangan, meskipun skor CVSS-nya masih berada pada tingkat tinggi, yaitu sekitar 7,47.

Tabel 3. Analisis sensitivitas ekonomi berdasarkan skenario asumsi implementasi GenAI

Komponen	Tradisional (Rp/tahun)	GenAI (Rp/tahun)	Selisih/dampak	Status
Biaya tenaga kerja	176.640.000	34.560.000	Penghematan 142.080.000	Asumsi berbasis waktu proses
Biaya tools dasar	25.000.000	25.000.000	0	Asumsi biaya
Biaya operasional GenAI	–	200.000.000	Tambahan 200.000.000	Asumsi biaya
Total biaya langsung	201.640.000	259.560.000	Tambahan 57.920.000	Hasil perhitungan
Baseline ALE	700.000.000	–	–	Asumsi risiko
ALE setelah GenAI	–	574.000.000	Reduksi 126.000.000	Skenario reduksi 18%
Manfaat bersih	–	–	68.080.000	Hasil perhitungan
Incremental ROI	–	–	117,5%	ROI atas biaya tambahan
Benefit–cost ratio	–	–	2,18:1	Reduksi ALE dibagi biaya tambahan

Nilai *baseline* ALE, persentase reduksi ALE, biaya tools, dan biaya operasional GenAI merupakan parameter skenario apabila tidak didukung oleh catatan biaya dan insiden aktual. Oleh karena itu, hasil ROI tidak ditafsirkan sebagai return aktual, melainkan estimasi sensitivitas ekonomi.

Dibandingkan pendekatan tradisional (NIST 800-30/FAIR manual) pada organisasi pembanding, waktu prioritisasi risiko dalam studi ini diposisikan sebagai penghematan *effort* analis keamanan: rata-rata dari 18,4 hari kerja/siklus menjadi 3,6 hari kerja/siklus. Berdasarkan Tabel 3, untuk menjaga konsistensi skala perguruan tinggi kecil, evaluasi efisiensi dihitung dengan asumsi operasional 4 siklus assessment per tahun dan keterlibatan 2 personel TI/keamanan. Dengan biaya tenaga kerja *fully loaded* yang konservatif untuk institusi kecil sebesar Rp1.200.000 per orang per hari, maka biaya tenaga kerja per tahun pada pendekatan tradisional adalah: $18{,}4 \times 4 \times 2 \times$ Rp1.200.000 = Rp176.640.000. Pada pendekatan GenAI, biaya tenaga kerja per tahun menjadi: $3{,}6 \times 4 \times 2 \times$ Rp1.200.000 = Rp34.560.000, sehingga terjadi penghematan biaya tenaga kerja sebesar Rp142.080.000 per tahun. Hasil ini menunjukkan peningkatan efisiensi pemrosesan risiko berbasis otomasi AI sebagaimana dilaporkan dalam studi pemanfaatan AI untuk cyber risk automation [9].

Untuk memodelkan *direct cost* yang lebih realistis pada perguruan tinggi kecil, pendekatan tradisional diasumsikan tetap membutuhkan perangkat pemindaian dan operasional minimum (misalnya scanner, logging dasar) sebesar Rp25.000.000 per tahun, sehingga biaya langsung tradisional (tenaga kerja + tools) = Rp176.640.000 + Rp25.000.000 = Rp201.640.000 per tahun. Pada skenario GenAI, selain biaya tools yang sama (Rp25.000.000), terdapat biaya operasional *GenAI Risk Engine* (komputasi, penyimpanan log, pemeliharaan, dan MLOps minimum) yang lebih masuk akal untuk institusi kecil tetapi tetap signifikan, misalnya Rp200.000.000 per tahun. Dengan demikian, biaya langsung GenAI = Rp34.560.000 + Rp25.000.000 + Rp200.000.000 = Rp259.560.000 per tahun. Artinya, GenAI membutuhkan investasi tambahan (incremental direct cost) sebesar Rp259.560.000 − Rp201.640.000 = Rp57.920.000 per tahun.

Komponen manfaat terbesar pada organisasi kecil umumnya bukan sekadar efisiensi waktu, melainkan penurunan ekspektasi kerugian tahunan (ALE) akibat penanganan kerentanan yang lebih cepat dan prioritisasi yang lebih tepat. Untuk skala perguruan tinggi kecil, nilai ALE baseline yang konservatif dan logis dapat diposisikan pada Rp700.000.000 per tahun (gabungan risiko downtime layanan akademik, pemulihan insiden, dan potensi dampak kebocoran data pada skala lokal). Dengan GenAI, penurunan ALE tidak diasumsikan ekstrem; digunakan penurunan moderat 18% (karena pengujian penetrasi terbatas dan implementasi awal), sehingga ALE baru ≈ 82% × Rp700.000.000 = Rp574.000.000 per tahun. Maka reduksi kerugian (loss reduction) = Rp700.000.000 − Rp574.000.000 = Rp126.000.000 per tahun. Model kuantifikasi ini mengikuti kerangka FAIR yang memungkinkan estimasi risiko finansial berbasis frekuensi dan dampak [12].

Berdasarkan komponen di atas, manfaat bersih tahunan dapat dihitung sebagai: (reduksi ALE) − (biaya tambahan GenAI) = Rp126.000.000 − Rp57.920.000 = Rp68.080.000 per tahun. Dengan definisi ROI berbasis investasi tambahan (*incremental ROI*): ROI = manfaat bersih / biaya tambahan = Rp68.080.000 / Rp57.920.000 = 1,175 atau 117,5% per tahun. Sementara itu, *benefit–cost ratio* berbasis reduksi kerugian saja adalah: Rp126.000.000 / Rp57.920.000 = 2,18, yang berarti setiap Rp1 investasi tambahan GenAI menghasilkan sekitar Rp2,18 reduksi ekspektasi kerugian tahunan pada konteks perguruan tinggi kecil dalam skenario ini. Pendekatan kuantifikasi ROI ini mengacu pada model evaluasi investasi keamanan informasi berbasis pengurangan eksposur risiko [26].

Copyright (c) HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi

Jika skenario yang sama dipaksakan pada UKM, ROI bisa menjadi lebih rendah atau bahkan negatif, karena ALE UKM umumnya jauh lebih kecil daripada perguruan tinggi (misalnya baseline ALE Rp150.000.000–Rp300.000.000/tahun). Dengan baseline ALE Rp200.000.000 dan penurunan 18% (Rp36.000.000), sementara biaya tambahan GenAI tetap ~Rp57.920.000, maka manfaat bersih menjadi negatif. Ini menjelaskan secara logis mengapa GenAI risk assessment lebih ekonomis untuk institusi yang memiliki eksposur risiko dan kompleksitas lebih tinggi, sedangkan pada UKM implementasi idealnya berbentuk shared service, *managed service*, atau versi lebih ringan agar biaya tambahan turun.

Hasil Validasi *Penetration Testing* Terbatas

Penetration testing terbatas dilakukan pada 10 *website* dengan lingkup non-destruktif dan terkontrol. Dari 1.284 skenario, 146 skenario (11,3%) tervalidasi secara empiris sebagai jalur eksploitasi realistis dalam lingkungan uji. Skenario dikategorikan sebagai "tervalidasi secara empiris" hanya apabila seluruh kriteria utama, yaitu keberadaan aset, kesesuaian kerentanan, reproduksibilitas, bukti teknis, dan konfirmasi, terpenuhi. Skenario yang hanya memiliki kesesuaian naratif atau hasil pemindaian otomatis tanpa verifikasi manual tidak dikategorikan sebagai tervalidasi secara empiris. Stratifikasi dilakukan berdasarkan jenis aset, kategori kerentanan, tingkat prioritas risiko, tipe jalur serangan, dan teknik MITRE ATT&CK. Pemilihan tidak didasarkan semata-mata pada skor tertinggi karena tujuan validasi adalah menilai keterwakilan beberapa kategori skenario, bukan hanya mengonfirmasi skenario yang paling berisiko. Setiap strata memiliki jumlah minimum skenario yang ditentukan sebelum pengujian. Skenario dengan risiko tinggi dan dampak potensial terhadap autentikasi, API, basis data, dan kontrol akses memperoleh prioritas validasi, sedangkan skenario berisiko rendah tetap diikutsertakan secara terbatas sebagai pembanding. Sebanyak 74% temuan penetrasi berkaitan dengan *improper access control* dan *insecure API endpoint*, sedangkan 19% terkait *outdated dependency libraries*. Validasi empiris terhadap attack graph memperlihatkan kesesuaian struktur jalur serangan yang diprediksi oleh simulator dengan konfigurasi sistem aktual. Prinsip pengujian terbatas dan responsible disclosure mengikuti praktik etis pengujian keamanan pada institusi akademik [18].

Validasi dilakukan pada lingkungan staging atau lingkungan uji dengan pengujian menggunakan pendekatan non-destruktif dan tidak mencakup penghapusan data, perubahan konfigurasi permanen, denial-of-service, privilege escalation yang tidak diperlukan, atau pengambilan data pribadi. Tahapan validasi meliputi: (1) verifikasi keberadaan aset dan layanan; (2) verifikasi versi atau konfigurasi yang berkaitan dengan kerentanan; (3) pemeriksaan prasyarat eksploitasi; (4) pengujian terbatas terhadap indikasi kerentanan; (5) pencatatan bukti teknis; dan (6) konfirmasi hasil oleh penguji keamanan atau pengelola sistem.

Hasil Kualitatif: Validasi Pemangku Kepentingan

Tahap kualitatif melibatkan wawancara semi-terstruktur terhadap 10 pengelola TI dan pimpinan unit sistem informasi dari 10 perguruan tinggi. Analisis tematik menghasilkan empat tema utama:

Tema 1: peningkatan visibilitas risiko yang mana responden menyatakan bahwa visualisasi *dynamic risk heatmap* mempermudah identifikasi prioritas mitigasi dibandingkan laporan statis berbasis daftar kerentanan.

Tema 2: efisiensi proses pengambilan Keputusan dibuktikan oleh sebanyak 80% responden yang menyatakan bahwa rekomendasi prioritas berbasis GenAI mempercepat proses alokasi sumber daya keamanan.

Tema 3: kekhawatiran transparansi algoritmik yang ditunjukkan oleh fakta sebagian responden (30%) mengemukakan kebutuhan dokumentasi transparan atas mekanisme perhitungan skor risiko oleh AI, sejalan dengan prinsip akuntabilitas dalam tata kelola AI [11].

Tema 4: kesiapan implementasi berkelanjutan yang mana responden mengidentifikasi kebutuhan pelatihan teknis untuk mengoperasikan sistem berbasis AI secara mandiri.

Integrasi temuan kuantitatif dan kualitatif menunjukkan konsistensi antara metrik risiko terukur dan persepsi pengguna terhadap kegunaan sistem.

Pembahasan

Hasil penelitian ini menegaskan bahwa implementasi *GenAI Multi-Agent Risk Engine* mampu mensimulasikan lebih dari 1.000 skenario serangan dan memperlihatkan penurunan waktu prioritisasi sebesar 14,8 hari dibandingkan pendekatan tradisional berbasis NIST 800-30 dan FAIR. Temuan kuantitatif menunjukkan pengurangan waktu prioritisasi sebesar lebih dari 80% serta peningkatan granularitas identifikasi jalur serangan melalui pemodelan graf berbobot. Hasil ini secara langsung menjawab rumusan masalah pertama dan kedua, yaitu kemampuan GenAI dalam menghasilkan skenario serangan komprehensif serta optimalisasi *automated/dynamic risk scoring*. Selain itu, analisis skenario menunjukkan incremental ROI sebesar 117,5% dan benefit–cost ratio sebesar 2,18:1 pada baseline ALE Rp700.000.000, reduksi ALE 18%, dan biaya tambahan GenAI Rp57.920.000 per tahun. Nilai tersebut merupakan estimasi berbasis asumsi, bukan ROI aktual yang telah diamati secara longitudinal. *Sensitivity* analysis menunjukkan bahwa kesimpulan ekonomi berubah secara material ketika persentase reduksi ALE atau biaya operasional GenAI berubah. Dalam konteks evaluasi investasi keamanan informasi, pendekatan berbasis pengurangan *annualized loss expectancy* (ALE) merupakan indikator yang diakui dalam literatur manajemen risiko siber [27]. Dengan demikian, penelitian ini menunjukkan bahwa otomasi berbasis AI dapat memperkuat efektivitas pengambilan keputusan strategis dalam tata kelola risiko siber pendidikan tinggi.

Copyright (c) HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi

Secara teoritis, integrasi model FAIR dengan simulasi generatif dan analisis graf memperluas pendekatan kuantifikasi risiko dari yang semula statis menjadi dinamis dan prediktif. FAIR secara tradisional mengasumsikan estimasi frekuensi dan dampak berdasarkan data historis dan *expert judgment* [12], sementara dalam penelitian ini nilai LEF diperkaya dengan probabilitas eksploitasi berbasis EPSS dan bobot centrality graf. Integrasi ini sejalan dengan perkembangan pendekatan *risk modeling* berbasis *data-driven analytics* yang menekankan pentingnya *probabilistic reasoning* dalam pengukuran risiko siber [28]. Penggunaan LLM untuk menghasilkan narasi risiko berbasis *chain-of-thought* juga memperluas dimensi *interpretatif risk communication*, mendukung tahap *Communicate* dalam NIST SP 800-30 melalui otomatisasi penyusunan laporan risiko yang terstruktur dan kontekstual [29]. Dengan demikian, penelitian ini mengoperasionalkan prinsip NIST dalam kerangka AI-*augmented risk lifecycle* yang lebih adaptif.

Dibandingkan dengan studi terdahulu yang memanfaatkan machine learning untuk deteksi intrusi atau klasifikasi kerentanan, penelitian ini memperluas cakupan penggunaan AI hingga tahap risk quantification dan remediation prioritization. Sebagian besar penelitian sebelumnya berfokus pada peningkatan akurasi deteksi serangan menggunakan deep learning tanpa mengintegrasikannya dengan model manajemen risiko formal [13]. Studi lain menunjukkan bahwa penggunaan AI dalam *cybersecurity* sering terfragmentasi antara intrusion detection dan vulnerability scanning tanpa pendekatan holistik berbasis *enterprise risk management* [30]. Dalam konteks pendidikan tinggi, penelitian mengenai kesiapan keamanan siber umumnya masih bersifat deskriptif dan belum memanfaatkan simulasi ancaman sintetis skala besar [31]. Oleh karena itu, hasil penelitian ini mengisi celah literatur dengan menawarkan model integratif yang menyatukan generative attack simulation, probabilistic risk modeling, dan evaluasi ekonomi.

Kontribusi ilmiah utama artikel ini terletak pada pengembangan arsitektur *GenAI-enhanced cyber risk assessment framework* yang mengintegrasikan tiga dimensi: (1) simulasi ancaman sintetis berbasis multi-agent generatif, (2) kuantifikasi risiko finansial menggunakan FAIR yang diperluas, dan (3) validasi empiris melalui penetration testing terbatas dan analisis kualitatif pemangku kepentingan. Kontribusi ini memperkaya diskursus mengenai penerapan Generative AI dalam domain keamanan siber, yang selama ini lebih banyak dibahas dalam konteks threat intelligence automation daripada risk governance [32]. Selain itu, pendekatan graph-based weighted attack path yang digunakan dalam penelitian ini memperlihatkan bahwa centrality metrics dapat berfungsi sebagai variabel prioritisasi risiko yang relevan dalam sistem kompleks [33]. Dengan demikian, penelitian ini berkontribusi terhadap pengembangan teori risk propagation dalam jaringan digital serta praktik implementasi AI dalam tata kelola keamanan institusi pendidikan.

Meskipun demikian, penelitian ini memiliki sejumlah keterbatasan. Pertama, jumlah unit analisis terbatas pada 10 website perguruan tinggi di Kalimantan sehingga generalisasi temuan perlu dilakukan secara hati-hati. Konteks geografis dan karakteristik infrastruktur digital di Kalimantan tidak secara otomatis merepresentasikan seluruh perguruan tinggi di Indonesia. Kedua, skenario serangan sintetis bergantung pada kualitas data CVE, EPSS, asset inventory, dan konfigurasi model LLM. Literatur mengenai AI governance menekankan bahwa sistem berbasis LLM rentan terhadap bias data dan interpretasi yang tidak sepenuhnya transparan [34]. Ketiga, penetration testing dilakukan dalam lingkup terbatas dan tidak mencakup seluruh variasi konfigurasi sistem produksi. Selain itu, evaluasi ROI dilakukan dalam horizon satu tahun sehingga belum menggambarkan dampak jangka panjang investasi keamanan berbasis AI.

Implikasi penelitian ini bagi praktisi adalah perlunya integrasi AI generatif dalam siklus manajemen risiko yang berkelanjutan, bukan sekadar sebagai alat deteksi insiden. Institusi pendidikan tinggi dapat memanfaatkan model ini untuk meningkatkan visibilitas risiko dan efisiensi alokasi sumber daya keamanan. Bagi pembuat kebijakan, hasil ini menunjukkan pentingnya penyusunan pedoman tata kelola AI dalam keamanan siber sektor pendidikan, sejalan dengan rekomendasi global mengenai penguatan *AI-enabled cybersecurity governance* [35]. Untuk penelitian lanjutan, diperlukan pengujian model pada skala institusi yang lebih luas serta eksplorasi integrasi teknik reinforcement learning untuk adaptasi otomatis terhadap perubahan lanskap ancaman. Penelitian mendatang juga dapat mengkaji aspek *explainable* AI dalam konteks *risk scoring* untuk meningkatkan transparansi algoritmik dan kepercayaan pemangku kepentingan [36].

4. SIMPULAN

Penelitian ini mengembangkan *GenAI-Enhanced Cyber Risk Assessment Framework* untuk *website* perguruan tinggi di Kalimantan melalui integrasi Generator Agent, Evaluator Agent, dan Reasoning Agent. Sistem menghasilkan *synthetic scenarios* berdasarkan kombinasi aset, *vulnerability record*, konfigurasi, teknik MITRE ATT&CK, dan relasi *attack graph*. *Synthetic scenario* diperlakukan sebagai unit keluaran generatif, sedangkan CVE atau *vulnerability, edge*, dan *attack path* dipisahkan sebagai unit analisis yang berbeda.

Sebanyak 1.284 *synthetic scenarios* dihasilkan setelah proses validasi format, pemeriksaan bukti, dan deduplikasi. Sebanyak 146 skenario dipilih untuk penetration testing terbatas. Namun, hanya skenario yang memenuhi kriteria kesesuaian aset, reproduksibilitas, bukti teknis, dan konfirmasi yang dapat disebut tervalidasi secara empiris. Dengan demikian, hasil validasi tidak boleh dipahami sebagai pembuktian langsung terhadap seluruh 1.284 skenario.

Skor risiko dibentuk secara deterministik melalui kombinasi CVSS yang dinormalisasi, EPSS, dan *centrality score* dengan bobot dasar masing-masing 0,50, 0,30, dan 0,20. Bobot tersebut merepresentasikan prioritas pada

Copyright (c) HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi

tingkat keparahan teknis, probabilitas eksploitasi, dan posisi aset dalam jalur serangan. Robustness skor perlu dibaca bersama sensitivity analysis karena perubahan bobot dapat mengubah urutan prioritas mitigasi.

Berdasarkan skenario asumsi dengan baseline ALE Rp700.000.000, reduksi ALE 18%, dan biaya tambahan Rp57.920.000 per tahun, estimasi incremental ROI adalah 117,5% dan benefit–cost ratio sebesar 2,18:1. Kedua ukuran tersebut berbeda secara definisi. Karena baseline ALE, persentase reduksi ALE, dan biaya operasional GenAI masih berbasis parameter skenario, hasil ROI belum dapat diperlakukan sebagai return aktual yang telah terobservasi.

Penelitian ini terbatas pada 10 *website* perguruan tinggi di Kalimantan, penggunaan skenario sintetis, dan penetration testing dalam lingkup terbatas. Penelitian berikutnya perlu mendokumentasikan model LLM, *prompt*, parameter inferensi, *log* generasi, distribusi skenario, bukti validasi, serta data biaya dan insiden aktual agar hasil dapat direplikasi dan generalisasi dapat dilakukan secara lebih meyakinkan.

DAFTAR PUSTAKA

[1] F. M. H. Tjiptabudi and R. I. Ndaumanu, "Evaluasi Celah Keamanan Website Dana Pensiun X Melalui Penetration Testing Berdasarkan ISSAF Framework," *Jurnal Algoritma*, vol. 21, no. 2, pp. 9–17, Nov. 2024, doi: 10.33364/algoritma/v.21-2.1644.

[2] N. Rani, B. Saha, and S. K. Shukla, "A Comprehensive Survey of Automated Advanced Persistent Threat Attribution: Taxonomy, Methods, Challenges and Open Research Problems," *Journal of Information Security and Applications*, vol. 92, p. 104076, Jul. 2025, doi: 10.1016/j.jisa.2025.104076.

[3] F. Fotis, "Economic Impact of Cyber Attacks and Effective Cyber Risk Management Strategies: A Light Literature Review and Case Study Analysis," *Procedia Comput. Sci.*, vol. 251, pp. 471–478, 2024, doi: 10.1016/j.procs.2024.11.135.

[4] A. Tiwari, S. Singhal, and A. Srivastava, "Cybersecurity Data Science From a Machine Learning Perspective," *OPSEARCH*, Oct. 2025, doi: 10.1007/s12597-025-01035-4.

[5] R. Belen-Saglam, H. Yuan, M. S. Heering, R. Ashraf, and S. Li, "A Systematic Literature Review on Cyber Security and Privacy Risks in MaaS (Mobility-as-a-Service) Systems," *Information*, vol. 16, no. 7, p. 514, Jun. 2025, doi: 10.3390/info16070514.

[6] H. Netshiunda and S. Madzvamuse, "A Systematic Literature Review of Cybersecurity Awareness and Strategy in Rural-Based Universities," *International Journal of Applied Research in Business and Management*, vol. 6, no. 5, Dec. 2025, doi: 10.51137/wrp.ijarbm.350.

[7] S. Islam, N. Basheer, S. Papastergiou, M. Ciampi, and S. Silvestri, "Intelligent Dynamic Cybersecurity Risk Management Framework with Explainability and Interpretability of AI Models for Enhancing Security and Resilience of Digital Infrastructure," *J. Reliab. Intell. Environ.*, vol. 11, no. 3, p. 12, Sep. 2025, doi: 10.1007/s40860-025-00253-3.

[8] A. Gain and M. S. Barik, "Attack Graph Based Security Metrics for Dynamic Networks," in *Information Systems Security: 19th International Conference*, Raipur, India, Dec. 2023, pp. 109–128. doi: 10.1007/978-3-031-49099-6_7.

[9] P. Balasubramanian *et al.*, "Generative AI for Cyber Threat Intelligence: Applications, Challenges, and Analysis of Real-World Case Studies," *Artificial Intelligence Review*, vol. 58, no. 11, p. 336, Aug. 2025, doi: 10.1007/s10462-025-11338-z.

[10] H. U. Khan, R. A. Khan, H. S. Alwageed, A. O. Almagrabi, S. Ayouni, and M. Maddeh, "AI-driven Cybersecurity Framework for Software Development Based on the ANN-ISM Paradigm," *Sci. Rep.*, vol. 15, no. 1, p. 13423, Apr. 2025, doi: 10.1038/s41598-025-97204-y.

[11] E. Tabassi, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," Jan. 2023. doi: 10.6028/NIST.AI.100-1.

[12] Jack. Freund and Jack. Jones, *Measuring and Managing Information Risk : a FAIR Approach*. Butterworth-Heinemann, 2025.

[13] M. A. Ferrag *et al.*, "Generative AI in Cybersecurity: A Comprehensive Review of LLM Applications and Vulnerabilities," *Internet of Things and Cyber-Physical Systems*, vol. 5, pp. 1–46, 2025, doi: 10.1016/j.iotcps.2025.01.001.

[14] N. O. Jaffal, M. Alkhanafseh, and D. Mohaisen, "Large Language Models in Cybersecurity: A Survey of Applications, Vulnerabilities, and Defense Techniques," *AI*, vol. 6, no. 9, p. 216, Sep. 2025, doi: 10.3390/ai6090216.

[15] M. N.-E. Saulaiman, M. Kozlovszky, and A. Csilling, "Graph-Based Automation of Threat Analysis and Risk Assessment for Automotive Security," *Information*, vol. 16, no. 6, p. 449, May 2025, doi: 10.3390/info16060449.

[16] P. Cheimonidis and K. Rantos, "A Dynamic Risk Assessment and Mitigation Model," *Applied Sciences*, vol. 15, no. 4, p. 2171, Feb. 2025, doi: 10.3390/app15042171.

[17] Y. Hassan *et al.*, "Exploring the Mediating Role of Information Security Culture in Enhancing Sustainable Practices Through Integrated Systems Infrastructure," *Sustainability*, vol. 17, no. 2, p. 687, Jan. 2025, doi: 10.3390/su17020687.

Copyright (c) HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi

[18] M. Alhamed and M. M. H. Rahman, "A Systematic Literature Review on Penetration Testing in Networks: Future Research Directions," *Applied Sciences*, vol. 13, no. 12, p. 6986, Jun. 2023, doi: 10.3390/app13126986.
[19] C. Hongling *et al.*, "Mixed Methods Research in General Practice: Key Points of Design and Implementation and Case Analysis," *Chinese General Practice Journal*, vol. 2, no. 3, p. 100070, Sep. 2025, doi: 10.1016/j.cgpj.2025.100070.
[20] J. Akoka, I. Comyn-Wattiau, N. Prat, and V. C. Storey, "Knowledge Contributions in Design Science Research: Paths of Knowledge Types," *Decision Support Systems*, vol. 166, p. 113898, Mar. 2023, doi: 10.1016/j.dss.2022.113898.
[21] L. Bracciale, S. Riozzi, G. Panico, E. Raso, G. Bianchi, and P. Loreti, "Quantifying Medical Device Cybersecurity Risk with CVSS BTE," *Scientific Reports*, vol. 15, no. 1, p. 42635, Nov. 2025, doi: 10.1038/s41598-025-26898-x.
[22] M. de F. Brilhante, D. Pestana, P. Pestana, and M. L. Rocha, "Measuring the Risk of Vulnerabilities Exploitation," *AppliedMath*, vol. 4, no. 1, pp. 20–54, Dec. 2023, doi: 10.3390/appliedmath4010002.
[23] M. Q. Li and B. C. M. Fung, "Security Concerns for Large Language Models: A survey," *Journal of Information Security and Applications*, vol. 95, p. 104284, Dec. 2025, doi: 10.1016/j.jisa.2025.104284.
[24] S. Unger, E. Arzoglou, M. Heinrich, D. Scheuermann, and S. Katzenbeisser, "Cybersecurity Risk Assessment in OT Systems Using Attack Graphs," *International Journal of Information Security*, vol. 25, no. 1, p. 34, Feb. 2026, doi: 10.1007/s10207-025-01198-7.
[25] Virginia. Braun and Victoria. Clarke, *Thematic Analysis : A Practical Guide*. SAGE, 2022.
[26] N. Y.-R. Douha, M. Sasabe, Y. Taenaka, and Y. Kadobayashi, "An Evolutionary Game Theoretic Analysis of Cybersecurity Investment Strategies for Smart-Home Users against Cyberattacks," *Applied Sciences*, vol. 13, no. 7, p. 4645, Apr. 2023, doi: 10.3390/app13074645.
[27] R. He, Z. Jin, and J. S.-H. Li, "Modeling and Management of Cyber Risk: A Cross-Disciplinary Review," *Annals of Actuarial Science*, vol. 18, no. 2, pp. 270–309, Jul. 2024, doi: 10.1017/S1748499523000258.
[28] R. Mohawesh, M. A. Ottom, and H. B. Salameh, "A Data-Driven Risk Assessment of Cybersecurity Challenges Posed by Generative AI," *Decision Analytics Journal*, vol. 15, p. 100580, Jun. 2025, doi: 10.1016/j.dajour.2025.100580.
[29] Y. Chen et al., "A survey of Large Language Models for Cyber Threat Detection," *Comput. Secur.*, vol. 145, p. 104016, Oct. 2024, doi: 10.1016/j.cose.2024.104016.
[30] M. Taddeo, T. McCutcheon, and L. Floridi, "Trusting Artificial Intelligence in Cybersecurity Is a Double-Edged Sword," 2022, pp. 289–297. doi: 10.1007/978-3-030-81907-1_15.
[31] O. Afolalu and M. S. Tsoeu, "Cybersecurity in Higher Education Institutions: A Systematic Review of Emerging Trends, Challenges and Solutions," *Future Internet*, vol. 17, no. 12, p. 575, Dec. 2025, doi: 10.3390/fi17120575.
[32] I. H. Sarker, H. Janicke, A. Mohsin, A. Gill, and L. Maglaras, "Explainable AI for Cybersecurity Automation, Intelligence and Trustworthiness in Digital Twin: Methods, Taxonomy, Challenges and Prospects," *ICT Express*, vol. 10, no. 4, pp. 935–958, Aug. 2024, doi: 10.1016/j.icte.2024.05.007.
[33] D. Frąszczak, "Identifying Network Propagation Sources Using Advanced Centrality Measures," *Entropy*, vol. 27, no. 9, p. 948, Sep. 2025, doi: 10.3390/e27090948.
[34] J. C. L. Ong *et al.*, "Ethical and Regulatory Challenges of Large Language Models in Medicine," *Lancet Digit. Health*, vol. 6, no. 6, pp. e428–e432, Jun. 2024, doi: 10.1016/S2589-7500(24)00061-X.
[35] R. Bhandari, "AI and Cybersecurity: Opportunities, Challenges, and Governance," *EDPACS*, pp. 1–9, Aug. 2025, doi: 10.1080/07366981.2025.2544363.
[36] T. Hulsen, "Explainable Artificial Intelligence (XAI): Concepts and Challenges in Healthcare," *AI*, vol. 4, no. 3, pp. 652–666, Aug. 2023, doi: 10.3390/ai4030034.

Copyright (c) HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi