PENGEMBANGAN *THREAT INTELLIGENCE KNOWLEDGE GRAPH* DENGAN *ENTITY EXTRACTION* TERHADAP *ADVANCED PERSISTENT THREAT* MENGGUNAKAN *PRE-TRAINED DEEPSEEK*

SKRIPSI

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

OLEH:

HABIB AL ASSYARI
09011382126142

JURUSAN SISTEM KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA
2025

HALAMAN PENGESAHAN

SKRIPSI

PENGEMBANGAN *THREAT INTELLIGENCE KNOWLEDGE GRAPH* DENGAN *ENTITY EXTRACTION* TERHADAP *ADVANCED PERSISTENT THREAT* MENGGUNAKAN *PRE-TRAINED DEEPSEEK*

Sebagai salah satu syarat untuk penyelesaian studi di

Program Studi S1 Sistem Komputer

Oleh:

HABIB AL ASSYARI

09011382126142

Pembimbing 1	**:**	**Prof. Ir. Deris Stiawan, M.T., Ph.D.**
		NIP. 197806172006041002
Pembimbing 2	**:**	**Nurul Afifah, M.Kom**
		NIP. 199211102025062006

Mengetahui

Ketua Jurusan Sistem Komputer

Dr. Ir. Sukemi, M.T

196612032006041001

AUTHENTICATION PAGE

THESIS

DEVELOPMENT OF THREAT INTELLIGENCE KNOWLEDGE GRAPH WITH ENTITY EXTRACTION FOR ADVANCED PERSISTENT THREATS USING PRE-TRAINED DEEPSEEK

As one of the requirements for the completion of undergraduate studies in the

Computer Systems Study Program

By:

HABIB AL ASSYARI

09011382126142

Supervisor	**:**	**Prof. Ir. Deris Stiawan, M.T., Ph.D.**
		NIP. 197806172006041002
Co-Supervisor	**:**	**Nurul Afifah, M.Kom**
		NIP. 199211102025062006

Acknowledged by

Head of Computer Systems Department

Dr. Ir. Sukemi, M.T

196612032006041001

LEMBAR PERSETUJUAN

Telah diuji pada:

Hari : Kamis

Tanggal : 24 Juli 2025

Tim Penguji:

1. **Ketua** : **Dr. Ir. Ahmad Heryanto, M.T.**
2. **Penguji** : **Yoppy Sazaki, S.Si., M.T.**
3. **Pembimbing I** : **Prof. Ir. Deris Stiawan, M.T.,Ph.D.**
4. **Pembimbing II** : **Nurul Afifah, M.Kom.**

Mengetahui, 24/9/25

Ketua Jurusan Sistem Komputer

Fakultas Ilmu Komputer

Universitas Sriwijaya

Dr. Ir. Sukemi M.T.

NIP. 196612032006041001

HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini:

Nama	:	Habib Al Assyari
NIM	:	09011382126142
Judul	:	Pengembangan *Threat Intelligence Knowledge Graph* Dengan *Entity Extraction* Terhadap *Advanced Persistent Threat* Menggunakan *Pre-Trained Deepseek*

Hasil Pengecekan Software Turnitin: 1%

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan atau plagiat. Apabila ditemukan unsur penjiplakan atau plagiat dalam laporan tugas akhir ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya.

Demikian, pernyataan ini saya buat dalam keadaan sadar tanpa paksaan dari siapapun.

Palembang, 24 September 2025

Habib Al Assyari

NIM. 09011382126142

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Segala puji dan syukur penulis panjatkan kepada Allah SWT atas limpahan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul "**Pengembangan *Threat Intelligence Knowledge Graph* dengan *Entity Extraction* terhadap *Advanced Persistent Threat* Menggunakan *Pre-Trained Deepseek***". Skripsi ini merupakan salah satu syarat untuk memenuhi mata kuliah skripsi pada jurusan Sistem Komputer, Universitas Sriwijaya.

Penyelesaian skripsi ini terwujud berkat bimbingan, dukungan, serta bantuan moril dan materil yang penulis peroleh dari berbagai pihak. Penulis menghaturkan ucapan terima kasih yang tulus dan sebesar-besarnya kepada:

1. Allah Subhanahu Wata'ala yang telah memberikan berkah serta nikmat kesehatan dan kesempatan kepada penulis untuk menyelesaikan studi dan Skripsi.
2. Bapak Prof. Dr. Taufiq Marwa, S.E., M.Si. selaku Rektor Universitas Sriwijaya.
3. Bapak Prof. Dr. Erwin, S.Si., M.Si. Selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Dr. Ir. Sukemi, M.T. Selaku Ketua Jurusan Sistem Komputer.
5. Bapak Huda Ubaya, S.T., M.T. Selaku Sekretaris Jurusan Sistem Komputer.
6. Bapak Prof. Dr. Ir. Bambang Tutuko, M.T. Selaku Dosen Pembimbing Akademik.
7. Bapak Prof. Ir. Deris Stiawan, M.T., Ph.D. Selaku Dosen Pembimbing I Tugas Akhir.
8. Ibu Nurul Afifah, M.Kom. Selaku Dosen Pembimbing II Tugas Akhir.
9. Kak Dr.(C). Dendi Renaldo Permana, S.Kom. Selaku Peneliti di Comnets Research Group yang membantu, membimbing, dan serta memberi arahan dalam penelitian.
10. Mbak Titi Fitriyanti Staf administrasi Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya Kampus Bukit Besar Palembang yang telah memberikan kemudahan dalam hal administrasi.

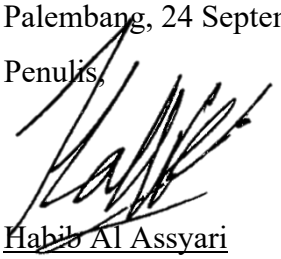
11. Kedua Orang Tua dan Keluarga yang selalu mendoakan serta memberikan motivasi dan semangat.
12. Teman-teman terutama dari “Grup Sedolor”, “Naks Rantau (Zootopia)” dan “yang aktif² be” yang selalu memberikan semangat dan dukungan selama menjalani studi.
13. Teman dan rekan seperjuangan Jurusan Sistem Komputer Angkatan 2021 terkhusus Sistem Komputer Unggulan 21.
14. Seluruh pihak yang tidak dapat penulis sebutkan satu persatu, yang telah banyak memberikan bantuan dan dukungan, penulis ucapkan terimakasih yang sebesar-besarnya.
15. Almamater Universitas Sriwijaya

Rasa terima kasih dan syukur yang tulus penulis sampaikan kepada berbagai pihak atas segala dukungan dan bantuannya, yang memungkinkan penelitian ini selesai tepat waktu. Semoga kebaikan tersebut menjadi ladang pahala, dan semoga kesuksesan serta kelancaran senantiasa menyertai semua pihak yang telah membantu. Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Untuk itu saran beserta kritikan yang membangun sangat diharapkan. Penulis berharap semoga skirpsi ini dapat bermanfaat bagi banyak orang khususnya bagi mahasiswa Fakultas Ilmu Komputer Universitas Sriwijaya secara langsung atau pun tidak langsung sebagai sumbangan pikiran dalam peningkatan mutu, semoga skripsi ini bermanfaat dan di ridhoi oleh Allah SWT. Semoga Allah SWT selalu melimpahkan rahmat dan karunia-Nya kepada kita semua.

Wassalamu’alaikum Warahamatullahi Wabarakatuh.

Palembang, 24 September 2025

Penulis,



Habib Al Assyari
NIM. 09011382126142

PENGEMBANGAN *THREAT INTELLIGENCE KNOWLEDGE GRAPH* DENGAN *ENTITY EXTRACTIO*N TERHADAP *ADVANCED PERSISTENT THREAT* MENGGUNAKAN *PRE-TRAINED DEEPSEEK*

HABIB AL ASSYARI (09011382126142)

Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya

Email: 09011382126142@student.unsri.ac.id

ABSTRAK

Penelitian ini bertujuan untuk mengatasi tantangan keamanan siber yang kompleks, yaitu *Advanced Persistent Threat* (APT), dengan mengembangkan *Threat Intelligence Knowledge Graph*. Penelitian mengusulkan pendekatan berbasis *Natural Language Processing* untuk mengekstrak entitas dan relasi dari laporan APT, menggunakan model *Deepseek* yang telah dilatih. Model ini di-*fine-tune* khusus untuk tugas *Named Entity Recognition* (NER) dan *Relation Extraction*. Hasil penelitian menunjukkan bahwa model *Deepseek* yang di-*fine-tune* mencapai hasil *f1-score* 0,960, mengungguli model BERT yang hanya memperoleh 0,694 pada dataset dan kondisi yang sama. Output utama dari penelitian ini adalah sebuah *knowledge graph* yang secara efektif memvisualisasikan entitas serangan, seperti *threat actor*, *malware*, dan taktik, ke dalam representasi terstruktur yang sesuai dengan standar *Structured Threat Information eXpression* (STIX). Temuan ini membuktikan bahwa *knowledge graph* dapat menjadi alat andal bagi analis keamanan dalam menganalisis pola serangan APT dengan lebih cepat dan mendalam.

Kata Kunci: *Cyber Threat Intelligence, Threat Intelligence Knowledge Graph, Entity Extraction, Advanced Persistent Threat, Deepseek, Named Entity Recognition, Relation Extraction, Structured Threat Information eXpression.*

DEVELOPMENT OF THREAT INTELLIGENCE KNOWLEDGE GRAPH WITH ENTITY EXTRACTION FOR ADVANCED PERSISTENT THREATS USING PRE-TRAINED DEEPSEEK

HABIB AL ASSYARI (09011382126142)

Department of Computer Systems, Faculty of Computer Science,
Sriwijaya University

Email: 09011382126142@student.unsri.ac.id

ABSTRACT

This research aims to address complex cybersecurity challenges, namely Advanced Persistent Threat (APT), by developing a Threat Intelligence Knowledge Graph. The study proposes a Natural Language Processing based approach to extract entities and relationships from APT reports, using a pre-trained Deepseek model. This model was fine-tuned specifically for Named Entity Recognition (NER) and Relation Extraction tasks. The research findings show that the fine-tuned Deepseek model achieved an F1-score of 0.960, outperforming the BERT model, which only achieved 0.694 under the same conditions and dataset. The primary output of this research is a knowledge graph that effectively visualizes attack entities, such as threat actors, malware, and tactics, into a structured representation that complies with the Structured Threat Information eXpression (STIX) standard. These findings demonstrate that a knowledge graph can be a reliable tool for security analysts to analyze APT attack patterns more quickly and in-depth.

Keywords: *Cyber Threat Intelligence, Threat Intelligence Knowledge Graph, Entity Extraction, Advanced Persistent Threat, Deepseek, Named Entity Recognition, Relation Extraction, Structured Threat Information eXpression*.

DAFTAR ISI

Halaman

DAFTAR GAMBAR

Halaman

DAFTAR TABEL

Halaman

BAB I
PENDAHULUAN

1.1 Latar Belakang

Cyber threat intelligence (CTI) merupakan sebuah sistem untuk mengumpulkan, memproses, dan menganalisis informasi mengenai ancaman siber yang dihadapi organisasi. CTI berfungsi sebagai sistem peringatan dini terhadap kejahatan digital. Sistem ini membantu organisasi memahami ancaman, mengidentifikasi potensi serangan, dan mengambil langkah proaktif untuk melindungi aset penting [1]. Seiring meningkatnya kompleksitas ancaman siber, era digital ini memerlukan pendekatan inovatif untuk meningkatkan deteksi dan prediksi. Salah satu pendekatan yang menjanjikan adalah *Named Entity Recognition* (NER), sebuah teknik dalam *Natural Language Processing* (NLP) yang mampu mengekstraksi informasi kunci secara otomatis. Penerapan NER pada laporan CTI terbukti efektif dalam mengidentifikasi entitas penting, sehingga memungkinkan analisis dan respons yang lebih cepat serta terarah.[2].

Potensi NLP dalam meningkatkan efektivitas CTI telah ditunjukkan melalui berbagai aplikasi, mulai dari *data mining* hingga analisis otomatis [3]. Beberapa penelitian misalnya [4] menerapkan NLP untuk mengumpulkan intelijen dari sumber informal seperti media sosial atau menganalisis sentimen di forum publik. Pendekatan ini memang bermanfaat, hanya saja tantangan utama dalam operasional CTI tetap terletak pada pengolahan volume besar laporan teknis tidak terstruktur, seperti laporan *advanced persistent threat* (APT) [5].

Penerapan teknik NLP diusulkan menggunakan pendekatan NER sebagai *entity extraction* untuk mengurai laporan-laporan ancaman siber. Proses ini pertama-tama mengidentifikasi entitas-entitas krusial secara otomatis, seperti pelaku ancaman, *malware*, dan infrastruktur. Setiap entitas tersebut lalu diklasifikasikan ke dalam kategori yang sesuai dengan standar *Structured Threat Information eXpression* (STIX) dan mengekstrak informasi tersebut menjadi format terstruktur yang siap digunakan sehingga bisa menciptakan *threat intelligence knowledge graph* (TIKG) [6].

Transformasi data tidak terstruktur menjadi informasi terstruktur merupakan fungsi utama dari teknik *entity extraction* [7]. Teknik ini secara langsung menargetkan "siapa, apa, di mana, dan bagaimana" dari sebuah narasi ancaman. Informasi terstruktur berupa entitas yang teridentifikasi dan relasi yang menghubungkannya menjadi bahan baku esensial untuk membangun sebuah TIKG. TIKG sendiri adalah visualisasi dari hasil proses *entity* dan *relation extraction* yang sistematis [8].

Pengembangan TIKG dari berbagai sumber informasi telah menjadi arah riset yang utama untuk menstrukturkan data CTI [6]. Kerangka kerja yang efektif umumnya melibatkan dua subtugas NLP krusial yaitu *entity extraction untuk* mengidentifikasi objek-objek penting dan *relation extraction untuk memahami* bagaimana objek-objek tersebut saling terhubung. [9]. Kemajuan *large language models* (LLMs) menawarkan potensi besar untuk meningkatkan akurasi dan otomatisasi kedua tugas tersebut [10][11]

Beberapa penelitian relevan telah menjadi landasan bagi pendekatan ini. Dalam studi [2] mengusulkan sebuah pendekatan *hybrid* (*rule-based, lexicon-based*, dan *machine learning* CRF). Hasilnya menunjukkan pendekatan ini yang efektif untuk domain spesifik dengan *f1-score* mencapai 90.5%. Kelebihannya adalah performa yang solid tanpa memerlukan komputasi *deep learning* yang intensif. Kekurangannya terletak pada ketergantungan yang tinggi pada rekayasa fitur dan aturan manual, sehingga kurang adaptif terhadap entitas ancaman jenis baru.

Riset [12] berfokus pada pemanfaatan melalui fine-tuning Llama-2, namun fokusnya lebih pada antarmuka percakapan (*chat*) untuk menjawab pertanyaan analitis, bukan pada pembangunan *knowledge graph* yang sistematis berbasis standar STIX. Laporan APT dipilih sebagai sumber data utama karena kekayaan informasi kontekstual dan teknis yang dimilikinya mengenai *campaign* serangan siber tingkat lanjut. Namun, laporan-laporan ini memiliki karakteristik bahasa yang padat, penuh dengan jargon teknis, dan menjelaskan hubungan antar entitas yang kompleks. Karakteristik tersebut membuat analisis manual menjadi lambat dan rentan terhadap kesalahan, sehingga menuntut penggunaan model NLP yang memiliki kapabilitas pemahaman konteks yang superior [13] .

Pemilihan model *pre-trained* Deepseek didasarkan pada arsitekturnya yang modern dan data pelatihannya yang unik. Berbeda dengan model LLM generasi sebelumnya seperti *Bidirectional Encoder Representations from Transformers* (BERT) atau *robustly optimized* BERT *pretraining approach* (RoBERTa) yang dominan dilatih pada teks bahasa alami, *Deepseek* dilatih pada korpus data masif yang tidak hanya mencakup bahasa alami, tetapi juga miliaran baris kode sumber. Kemampuan ganda ini memberinya keunggulan inheren dalam memahami teks yang terstruktur secara logis dan padat akan terminologi teknis, menjadikannya kandidat ideal untuk secara akurat mengurai kompleksitas laporan APT yang sering kali memadukan narasi dan elemen mirip kode. Kemampuan dasar inilah yang akan dispesialisasi lebih lanjut melalui proses *fine-tuning* [14] [15].

Paradigma *pre-trained* dan *fine-tuning*, yang dipopulerkan oleh model seperti BERT, memungkinkan adaptasi model pada domain yang spesifik [16]. STIX digunakan sebagai acuan representasi pengetahuan guna memastikan hasil ekstraksi memiliki interoperabilitas dan dapat digunakan secara luas [11]. Kinerja model NER akan dievaluasi secara cermat menggunakan metrik standar (*accuracy, precision, recall*, dan *f1-score*), sebagaimana yang umum dilakukan dalam studi NER pada domain keamanan siber [17].

Penelitian [18] berfokus pada pengembangan SecBERT, model berbasis BERT yang dilatih khusus pada korpus keamanan siber untuk mendukung analisis laporan ancaman. Dengan metode *pre-training* dan *fine-tuning* pada tugas seperti klasifikasi dan NER, SecBERT menunjukkan keunggulan dalam memahami terminologi teknis dibanding model NLP umum. Meski membutuhkan data yang beragam dan komputasi besar, hasil evaluasi membuktikan bahwa SecBERT melampaui model *baseline* dan berkontribusi pada peningkatan akurasi sistem CTI. SecBERT mencapai kinerja terbaik pada tugas NER dengan *F1-score* 90,6%, lebih tinggi dibandingkan model generik maupun domain-spesifik lainnya, sehingga menegaskan efektivitasnya dalam ekstraksi entitas keamanan siber.

Penelitian ini merancang kerangka kerja berbasis LLM *Deepseek* untuk mengekstrak entitas dari laporan APT dan membangun TIKG yang dinamis [7]. Pendekatan *fine-tuning* pada model pre-trained seperti BERT digunakan agar sesuai dengan domain keamanan siber [16], sementara STIX dijadikan acuan representasi

pengetahuan guna memastikan interoperabilitas hasil ekstraksi [11]. Evaluasi model dilakukan dengan metrik standar (*accuracy, precision, recall*, dan *F1-score*) sebagaimana umum digunakan dalam studi NER pada domain keamanan siber [17].

Pemilihan laporan APT sebagai sumber data utama didasarkan pada kekayaan informasi kontekstual dan teknis yang dimilikinya [19]. Laporan APT umumnya memuat detail mendalam mengenai *tactics, techniques, and procedures* (TTPs) yang digunakan oleh *threat actor*, serta hubungan antar entitas seperti *malware*, infrastruktur, dan *campaign*. Karakteristik ini menjadikan laporan APT sebagai sumber data yang relevan dan representatif untuk penelitian pengembangan TIKG, karena mampu mencerminkan kompleksitas ancaman dunia nyata [20].

Penggunaan deepseek dinilai lebih efektif dalam mengekstraksi entitas dalam suatu laporan APT dibandingkan *Pre-trained* model lainnya seperti BERT [21], RoBERTa [22], dan SecBERT [18] dengan harapan dapat mendukung pengembangan CTI yang lebih efektif dan responsif terhadap ancaman.berkontribusi secara signifikan dalam menciptakan sistem CTI yang lebih efektif dan mampu merespons ancaman siber yang semakin kompleks.

1.2 Rumusan Masalah

Rumusan masalah yang diangkat berdasarkan latar belakang pada penelitian ini adalah sebagai berikut.

1. Bagaimana merancang dan mengembangkan model ekstraksi entitas menggunakan *pre-trained Deepseek* untuk mengidentifikasi entitas-entitas kunci dari laporan APT berbahasa Inggris, dengan memanfaatkan *dataset* yang mencerminkan karakteristik laporan tersebut?
2. Bagaimana mengukur dan mengevaluasi performa model *Deepseek* dalam tugas *entity extraction* pada domain laporan APT menggunakan metrik evaluasi standar seperti *accuracy*, *precision*, *recall*, dan *f1-score*?
3. Bagaimana memvisualisasikan entitas dan potensi relasi yang diekstraksi dari laporan APT ke dalam sebuah TIKG dengan skema yang mengacu pada standar STIX untuk memastikan interoperabilitas dan pemahaman yang konsisten?

1.3 Batasan Masalah

Batasan masalah bertujuan agar penelitian ini lebih terfokus dan terarah. Beberapa batasan masalah yang ditetapkan adalah sebagai berikut.

1. Data yang digunakan dalam penelitian ini berupa laporan APT yang tersedia secara publik bersumber dari *open-source cyber threat intelligence* (OSCTI) *dataset* yaitu CyberNER *dataset* yang berjumlah 6.311 kalimat dan 203.989 kata.
2. Entitas yang diekstraksi berdasarkan aturan STIX *domain object* yang berjumlah 10 jenis entitas, yaitu *threat-actor, malware, tool, vulnerability, attack-pattern, indicator, identity, location, infrastructure,* dan *campaign*.
3. Metode yang digunakan pada penelitian ini, yaitu metode *Deepseek* yang dilakukan pemrosesan *fine tuning.*
4. Metode *Deepseek* dengan pendekatan NER menggunakan *library* atau *framework* NLP, yaitu *spaCy*, NLTK, atau *hugging face transformers*.
5. Evaluasi model akan dilakukan menggunakan metrik, seperti *accuracy*, *precision, recall*, dan *f1-score*.

1.4 Tujuan

Tujuan dari penulisan dan penelitian pada skripsi ini, yaitu sebagai berikut.

1. Merancang dan mengembangkan model ekstraksi entitas menggunakan *pre-trained Deepseek* untuk mengidentifikasi entitas-entitas kunci dari laporan APT berbahasa Inggris, dengan memanfaatkan *dataset* yang mencerminkan karakteristik laporan tersebut.
2. Mengukur dan mengevaluasi performa model *Deepseek* dalam tugas *entity extraction* pada domain laporan APT menggunakan metrik evaluasi standar seperti *accuracy*, *precision*, *recall*, dan *f1-score.*
3. Memvisualisasikan entitas dan potensi relasi yang diekstraksi dari laporan APT ke dalam sebuah TIKG dengan skema yang mengacu pada standar STIX untuk memastikan interoperabilitas dan pemahaman yang konsisten.

1.5 Manfaat

Diharapkan penelitian ini dapat memberikan manfaat sebagai berikut.

1. Bagi komunitas keamanan siber: Memberikan solusi otomatisasi untuk ekstraksi informasi penting dari laporan APT, sehingga mendukung analisis ancaman yang lebih cepat dan akurat.
2. Bagi akademisi dan peneliti: Menjadi referensi dalam pengembangan aplikasi NLP untuk keamanan siber.
3. Bagi industri: Mempermudah pengambilan keputusan strategis terkait mitigasi ancaman siber melalui informasi ancaman yang lebih terstruktur.

1.6 Sistematika Penulisan

Penulisan skripsi ini dituangkan ke dalam 5 (lima) bab yang dirancang untuk mempermudah pemahaman dan memperjelas isi setiap bab, yaitu sebagai berikut.

BAB I PENDAHULUAN

Merupakan bab pembuka yang memberikan gambaran umum mengenai penelitian. Di dalamnya akan diuraikan latar belakang penelitian yang menjelaskan konteks dan urgensi kenapa penelitian ini dilakukan, rumusan masalah yang mengidentifikasi pertanyaan-pertanyaan penelitian yang akan dijawab, batasan masalah untuk mempersempit fokus penelitian, tujuan penelitian yang ingin dicapai, manfaat penelitian baik secara teoretis maupun praktis, metodologi penelitian secara garis besar yang akan digunakan, dan diakhiri dengan sistematika penulisan yang menjelaskan struktur keseluruhan dokumen penelitian.

BAB II TINJAUAN PUSTAKA

Menyajikan tinjauan literatur yang relevan dan mendalam terkait dengan topik penelitian. Studi pustaka ini bertujuan untuk membangun landasan teoretis yang kuat, memahami *state-of-the-art* di bidang terkait, dan mengidentifikasi *research gap* yang akan diisi oleh penelitian ini. Bab ini akan mengkaji secara sistematis berbagai literatur, dimulai dari konsep fundamental seperti *Cyber Threat*

Intelligence (CTI) dan *Advanced Persistent Threat* (APT). Selanjutnya, akan dibahas landasan teoretis untuk hasil akhir yang dituju, yaitu *Knowledge Graph*, beserta standar representasi data CTI, *Structured Threat Information eXpression* (STIX). Pada tataran teknis, bab ini akan mengulas teknik-teknik *Natural Language Processing* (NLP) yang relevan, tidak hanya terbatas pada *Named Entity Recognition* (NER), tetapi juga mencakup Ekstraksi Relasi *(Relation Extraction)* yang krusial untuk pembangunan graf. Landasan teknologi akan dibahas secara hierarkis, mulai dari *Deep Learning*, kemudian sub-bidang spesifik *Large Language Models* (LLM), hingga justifikasi pemilihan model *pre-trained Deepseek*.

BAB III METODOLOGI PENELITIAN

Menjelaskan secara rinci metodologi penelitian yang digunakan untuk menjawab rumusan masalah dan mencapai tujuan penelitian. Bab ini akan menguraikan langkah-langkah penelitian secara sistematis, mulai dari pengumpulan data, pra-pemrosesan data, pengembangan model NER dengan fine-tuning *deepseek*, implementasi teknik NER untuk ekstraksi entitas STIX domain object, pengujian dan evaluasi model menggunakan metrik yang relevan, hingga visualisasi hasil ekstraksi entitas.

BAB IV HASIL DAN PEMBAHASAN

Menyajikan hasil penelitian yang diperoleh dari implementasi metodologi yang telah dijelaskan pada bab sebelumnya. Hasil penelitian akan disajikan secara sistematis dan jelas, diikuti dengan pembahasan yang mendalam. Pembahasan akan menginterpretasi hasil penelitian, menganalisis implikasi hasil terhadap rumusan masalah dan tujuan penelitian, membandingkan hasil penelitian dengan penelitian sebelumnya yang relevan (dari studi pustaka), serta mengidentifikasi potensi keterbatasan dan area pengembangan lebih lanjut.

BAB V KESIMPULAN DAN SARAN

Merupakan bab penutup yang menyajikan kesimpulan dari keseluruhan penelitian dan memberikan saran untuk penelitian selanjutnya. Kesimpulan akan merangkum temuan-temuan utama penelitian yang berhasil dicapai, menjawab rumusan masalah yang telah diajukan, dan menegaskan kontribusi penelitian. Saran akan berisi rekomendasi untuk penelitian selanjutnya, pengembangan lebih lanjut dari model atau kerangka kerja yang dihasilkan, serta potensi aplikasi praktis dari hasil penelitian.

DAFTAR PUSTAKA

[1] G. Sakellariou, P. Fouliras, I. Mavridis, and P. Sarigiannidis, "A Reference Model for Cyber Threat Intelligence (CTI) Systems," *Electronics (Switzerland)*, vol. 11, no. 9, pp. 1–23, 2022, doi: 10.3390/electronics11091401.

[2] H. Wu, X. Li, and Y. Gao, "An Effective Approach of Named Entity Recognition for Cyber Threat Intelligence," *Proceedings of 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference, ITNEC 2020*, no. Itnec, pp. 1370–1374, 2020, doi: 10.1109/ITNEC48623.2020.9085102.

[3] N. Sun *et al.*, "Cyber Threat Intelligence Mining for Proactive Cybersecurity Defense: A Survey and New Perspectives," *IEEE Communications Surveys and Tutorials*, vol. 25, no. 3, pp. 1748–1774, 2023, doi: 10.1109/COMST.2023.3273282.

[4] L. M. Kristiansen, V. Agarwal, K. Franke, and R. S. Shah, "CTI-Twitter: Gathering Cyber Threat Intelligence from Twitter using Integrated Supervised and Unsupervised Learning," *Proceedings - 2020 IEEE International Conference on Big Data, Big Data 2020*, pp. 2299–2308, 2020, doi: 10.1109/BigData50022.2020.9378393.

[5] N. Nurwanda, N. Suarna, and W. Prihartono, "Penerapan Nlp (Natural Language Processing) Dalam Analisis Sentimen Pengguna Telegram Di Playstore," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 2, pp. 1841–1846, 2024, doi: 10.36040/jati.v8i2.8469.

[6] G. Wang, P. Liu, J. Huang, H. Bin, X. Wang, and H. Zhu, "KnowCTI: Knowledge-based cyber threat intelligence entity and relation extraction," *Comput Secur*, vol. 141, no. March, 2024, doi: 10.1016/j.cose.2024.103824.

[7] T. Al-Moslmi, M. Gallofre Ocana, A. L. Opdahl, and C. Veres, "Named Entity Extraction for Knowledge Graphs: A Literature Overview," *IEEE Access*, vol. 8, pp. 32862–32881, 2020, doi: 10.1109/ACCESS.2020.2973928.

[8] ∗ and Gaosheng Wanga,b, Peipei Liua,b, Jintao Huanga,b, Haoyu Bina,b, Xi Wanga,b and B. Hongsong Zhua, "KnowCTI: Knowledge-based Cyber Threat Intelligence Entity and Relation Extraction," *Journal of Emerging Technologies and Innovative Research (JETIR)*, vol. 06, no. 4, pp. 7–20, 2020.

[9] F. Perrina, F. Marchiori, M. Conti, and N. V. Verde, “AGIR: Automating Cyber Threat Intelligence Reporting with Natural Language Generation,” *Proceedings - 2023 IEEE International Conference on Big Data, BigData 2023*, pp. 3053–3062, 2023, doi: 10.1109/BigData59044.2023.10386116.

[10] DeepSeek-AI *et al.*, “DeepSeek LLM: Scaling Open-Source Language Models with Longtermism,” 2024, [Online]. Available: http://arxiv.org/abs/2401.02954

[11] T. Sun, P. Yang, M. Li, and S. Liao, “An automatic generation approach of the cyber threat intelligence records based on multi-source information fusion,” *Future Internet*, vol. 13, no. 2, pp. 1–19, 2021, doi: 10.3390/fi13020040.

[12] L. E. Xiao, “LLM-DER : A Named Entity Recognition Method Based on Large Language Models for Chinese Coal Chemical Domain”.

[13] M. Tatam, B. Shanmugam, S. Azam, and K. Kannoorpatti, “A review of threat modelling approaches for APT-style attacks,” *Heliyon*, vol. 7, no. 1, p. e05969, 2021, doi: 10.1016/j.heliyon.2021.e05969.

[14] D. Guo *et al.*, “DeepSeek-Coder: When the Large Language Model Meets Programming -- The Rise of Code Intelligence,” pp. 1–23, 2024, [Online]. Available: http://arxiv.org/abs/2401.14196

[15] H. Xu *et al.*, “Large Language Models for Cyber Security: A Systematic Literature Review,” 2024, doi: 10.1145/3695988.

[16] J. Li, A. Sun, J. Han, and C. Li, “A Survey on Deep Learning for Named Entity Recognition,” *IEEE Trans Knowl Data Eng*, vol. 34, no. 1, pp. 50–70, 2022, doi: 10.1109/TKDE.2020.2981314.

[17] F. Marchiori, M. Conti, and N. V. Verde, “STIXnet: A Novel and Modular Solution for Extracting All STIX Objects in CTI Reports,” *ACM International Conference Proceeding Series*, 2023, doi: 10.1145/3600160.3600182.

[18] H. Huang and Y. Wang, “SecBERT: Privacy-preserving pre-training based neural network inference system,” *Neural Networks*, vol. 172, no. July 2023, p. 106135, 2024, doi: 10.1016/j.neunet.2024.106135.

[19] G. Husari, E. Al-Shaer, M. Ahmed, B. Chu, and X. Niu, “TTPDrill: Automatic and accurate extraction of threat actions from unstructured text of CTI Sources,” *ACM International Conference Proceeding Series*, vol. Part F1325, no. December, pp. 103–115, 2017, doi: 10.1145/3134600.3134646.

[20] X. Zhao, R. Jiang, Y. Han, A. Li, and Z. Peng, “A survey on cybersecurity knowledge graph construction,” *Comput Secur*, vol. 136, no. September 2023, p. 103524, 2024, doi: 10.1016/j.cose.2023.103524.

[21] M. Osaka, “Deep Learning-Powered Cyber Threat Intelligence for Multi-Cloud Environments Author : Amelia Ethan , Motohisa Osaka,” no. July 2024, 2025.

[22] B. Chen, H. Li, D. Zhao, Y. Yang, and C. Pan, “Quality assessment of cyber threat intelligence knowledge graph based on adaptive joining of embedding model,” *Complex and Intelligent Systems*, vol. 11, no. 1, pp. 1–14, 2025, doi: 10.1007/s40747-024-01661-3.

[23] S. Samtani, H. Chen, M. Kantarcioglu, and B. Thuraisingham, “Explainable Artificial Intelligence for Cyber Threat Intelligence (XAI-CTI),” *IEEE Trans Dependable Secure Comput*, vol. 19, no. 4, pp. 2149–2150, 2022, doi: 10.1109/TDSC.2022.3168187.

[24] Y. Lu, Y. Li, R. Zhang, W. Chen, B. Ai, and D. Niyato, “Graph Neural Networks for Wireless Networks: Graph Representation, Architecture and Evaluation,” *IEEE Wirel Commun*, vol. 32, no. 1, pp. 150–156, 2025, doi: 10.1109/MWC.006.2400131.

[25] I. MOUICHE and S. Saad, “Entity and Relation Extractions for Threat Intelligence Knowledge Graphs,” *Available at SSRN 4878306*, vol. 148, no. August 2024, 2025.

[26] L. M. Kristiansen, V. Agarwal, K. Franke, and R. S. Shah, “CTI-Twitter: Gathering Cyber Threat Intelligence from Twitter using Integrated Supervised and Unsupervised Learning,” *Proceedings - 2020 IEEE International Conference on Big Data, Big Data 2020*, pp. 2299–2308, 2020, doi: 10.1109/BigData50022.2020.9378393.

[27] M. F. Haque and R. Krishnan, “Toward Automated Cyber Defense with Secure Sharing of Structured Cyber Threat Intelligence,” *Information Systems Frontiers*, vol. 23, no. 4, pp. 883–896, 2021, doi: 10.1007/s10796-020-10103-7.

[28] G. Sakellariou, P. Fouliras, I. Mavridis, and P. Sarigiannidis, “A Reference Model for Cyber Threat Intelligence (CTI) Systems,” *Electronics (Switzerland)*, vol. 11, no. 9, pp. 1–23, 2022, doi: 10.3390/electronics11091401.

[29] N. Sun *et al.*, “Cyber Threat Intelligence Mining for Proactive Cybersecurity Defense: A Survey and New Perspectives,” *IEEE*

Communications Surveys and Tutorials, vol. 25, no. 3, pp. 1748–1774, 2023, doi: 10.1109/COMST.2023.3273282.

[30] A. R. Tapsoba and T. F. Ouédraogo, "A Relevant Feature Identification Approach to Detect APTs in HTTPS Traffic," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 9, pp. 969–977, 2024, doi: 10.14569/IJACSA.2024.0150999.

[31] D. R. Permana, D. Stiawan, D. P. Rini, N. Afifah, S. K. Ningrum, and R. Budiarto, "An Enhanced Method with Part of Speech Tagging and Named Entity Recognition Techniques Towards Advanced Persistent Threat in Cyber Threat Intelligence: Work in Progress," *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, no. September, pp. 493–498, 2024, doi: 10.1109/EECSI63442.2024.10776424.

[32] N. O'Brien, "Assessing the importance of modern security tools and frameworks to help detect and defend against Cozy bear," 2022.

[33] N. Mohamed, "State-of-the-Art in Chinese APT Attack and Using Threat Intelligence for Detection. A Survey," *Journal of Positive School Psychology*, vol. 2022, no. 5, pp. 4419–4443, 2022, [Online]. Available: http://journalppw.com

[34] P. Kálnai and M. Poslušný, "Lazarus Group: a mahjong game played with different sets of tiles," *Virus Bulletin International Conference*, no. September, 2018.

[35] F. K. Parast, "Equation Unsolved: Inside the Shadowy World of Elite Cyber Spies," in *2024 34th International Conference on Collaborative Advances in Software and COmputiNg (CASCON)*, 2024, pp. 1–6. doi: 10.1109/CASCON62161.2024.10838060.

[36] A. Dahan, "Operation cobalt kitty," *Cybereason. Retrieved from https://cdn2. hubspot. net/hubfs/3354902/Cybereason% 20Labs% 20Analysis% 20Operation% 20Cobalt% 20Kitty. pdf*, 2017.

[37] M. U. Rana, O. Ellahi, M. Alam, J. L. Webber, A. Mehbodniya, and S. Khan, "Offensive Security: Cyber Threat Intelligence Enrichment With Counterintelligence and Counterattack," *IEEE Access*, vol. 10, no. August, pp. 108760–108774, 2022, doi: 10.1109/ACCESS.2022.3213644.

[38] Y. Gao, X. Li, H. Peng, B. Fang, and P. S. Yu, "HinCTI: A Cyber Threat Intelligence Modeling and Identification System Based on Heterogeneous Information Network," *IEEE Trans Knowl Data Eng*, vol. 34, no. 2, pp. 708–722, 2022, doi: 10.1109/TKDE.2020.2987019.

[39] The MITRE Corporation, “About STIX.” Accessed: Jul. 15, 2025. [Online]. Available: https://stixproject.github.io/about/

[40] L. Roj and N. Lukač, “Integration of Named Entity Extraction Based on Deep Learning for Neo4j Graph Database,” pp. 11–14, 2024.

[41] T. Li, Y. Guo, and A. Ju, “A Self-Attention-Based Approach for Named Entity Recognition in Cybersecurity,” *Proceedings - 2019 15th International Conference on Computational Intelligence and Security, CIS 2019*, pp. 147–150, 2019, doi: 10.1109/CIS.2019.00039.

[42] N. Sciences, “Enhancing Cyber Threat Intelligence with STIX-Shifter : An Analysis of ACT and STIX Integration”.

[43] Z. X. Li, Y. J. Li, Y. W. Liu, C. Liu, and N. X. Zhou, “K-CTIAA: Automatic Analysis of Cyber Threat Intelligence Based on a Knowledge Graph,” *Symmetry (Basel)*, vol. 15, no. 2, 2023, doi: 10.3390/sym15020337.

[44] T. Shen, F. Zhang, and J. Cheng, “A comprehensive overview of knowledge graph completion,” *Knowl Based Syst*, vol. 255, p. 109597, 2022, doi: 10.1016/j.knosys.2022.109597.

[45] Nagendrababu NC, Samyama Gunjal GH, and Himabindhu N, “Advance persistent threat prediction using knowledge graph,” *International Journal of Science and Technology Research Archive*, vol. 6, no. 2, pp. 071–082, 2024, doi: 10.53771/ijstra.2024.6.2.0047.

[46] X. Wang, J. Yang, Q. Wang, and C. Su, “Threat intelligence relationship extraction based on distant supervision and reinforcement learning,” *Proceedings of the International Conference on Software Engineering and Knowledge Engineering, SEKE*, vol. PartF16244, pp. 572–576, 2020, doi: 10.18293/SEKE2020-149.

[47] N. Miloslavskaya, “Network Protection Tools for Network Security Intelligence Centers,” *Procedia Comput Sci*, vol. 190, pp. 597–603, 2021, doi: 10.1016/j.procs.2021.06.070.

[48] P. Alaeifar, S. Pal, Z. Jadidi, M. Hussain, and E. Foo, “Current approaches and future directions for Cyber Threat Intelligence sharing: A survey,” *Journal of Information Security and Applications*, vol. 83, no. May, 2024, doi: 10.1016/j.jisa.2024.103786.

[49] K. Li *et al.*, “InstructCoder: Instruction Tuning Large Language Models for Code Editing,” vol. 4, pp. 473–493, 2023, [Online]. Available: http://arxiv.org/abs/2310.20329

[50] P. C. Aravind, D. R. Arikkat, A. S. Krishnan, and B. Tesneem, *CyTIE : Cyber Threat Intelligence Extraction*, vol. 1. Springer Nature Switzerland. doi: 10.1007/978-3-031-59100-6.