BAB 1
PENDAHULUAN

1.1 Latar Belakang

Tingkat kompleksitas serangan phishing telah berkembang pesat seiring dengan kemajuan teknologi (Ajar Edi, Direktur Corporate Affairs Microsoft Indonesia, 2023), dan salah satu alat yang sangat populer dalam AI saat ini adalah GPT-4 (Generative Pre-trained Transformer 4). Dalam serangan phishing saat ini, penjahat siber dapat menggunakan GPT-4 untuk meningkatkan tingkat keberhasilan serangan mereka. GPT-4 menjanjikan kemajuan signifikan dalam bidang pemrosesan bahasa alami (Natural Language Processing/NLP) (Baktash et al, 2023). Model ini merupakan hasil pengembangan dari versi-versi sebelumnya dan telah mengalami perbaikan besar dalam pemahaman bahasa alami dan kemampuan generatifnya. GPT-4 dilatih dengan sejumlah besar data teks dari internet, sehingga bisa untuk menghasilkan teks yang sangat realistis dan serupa dengan tulisan manusia. Ini memiliki potensi besar dalam berbagai aplikasi NLP, seperti penerjemahan otomatis, analisis sentimen, dan pembuatan konten otomatis. Sehingga GPT-4 menghasilkan teks yang sangat realistis dan sulit dibedakan dari teks yang ditulis oleh manusia. Dengan kemampuan ini, GPT-4 dapat disalahgunakan untuk menciptakan pesan phishing yang sangat meyakinkan. Pesan tersebut dapat dibuat sedemikian rupa sehingga tampaknya berasal dari sumber yang sah, seperti perusahaan atau pemerintah, untuk memancing korban agar mengungkapkan informasi pribadi atau mengklik tautan berbahaya.

Penggunaan GPT-4 dalam upaya phishing membantu aktor jahat untuk membuat pesan menipu yang sulit dibedakan dari korespondensi/surat menyurat yang ditujukan untuk kepentingan bisnis asli. Dengan meniru organisasi terkemuka atau lembaga pemerintah, GPT-4 dapat menghasilkan teks yang mendorong korban untuk membocorkan data pribadi atau mengklik tautan berbahaya, meningkatkan risiko serangan yang berhasil dan membuatnya lebih menantang bagi sistem keamanan siber untuk mendeteksi ancaman tersebut karena kemampuan GPT-4 untuk menghasilkan pesan yang menyerupai komunikasi manusia. Hacker menggunakan kemampuan generatif dari alat GenAI (Generative AI) untuk membuat serangan rekayasa sosial yang meyakinkan berupa serangan phishing dan berbagai jenis kode berbahaya yang dapat dikompilasi menjadi sebuah berkas malware yang

dapat dijalankan (Gupta et al, 2023). Dengan menggunakan kecerdasan buatan ini, penyerang dapat membangun pesan rekayasa sosial yang meyakinkan, yang mungkin tampak berasal dari sumber yang sah atau teman yang dikenal, untuk memikat korban agar mengungkapkan informasi pribadi atau rahasia. Selain itu, serangan phishing juga menjadi lebih canggih karena alat GenAI dapat menghasilkan pesan yang tampaknya berasal dari lembaga keuangan, pemerintah, atau perusahaan yang dikenal. Pesan semacam ini dirancang untuk mengelabui korban agar mengklik tautan berbahaya atau mengungkapkan informasi seperti kata sandi atau nomor kartu kredit.

Artificial Intelligence (AI), yang dapat diterjemahkan sebagai Kecerdasan Buatan atau Kecerdasan Artifisial, terus menjadi teknologi yang semakin canggih pada berbagai bidang, terutama teknologi, perbankan, keuangan, hiburan, dan e-commerce. Disamping kebermanfaatannya, AI juga dapat memberikan ancaman pertahanan siber. Pertahanan siber telah menjadi salah satu prioritas utama bagi negara-negara di seluruh dunia. Ancaman siber yang semakin kompleks dan mengintai telah menuntut negara-negara untuk terus mengembangkan strategi dan teknologi baru dalam melindungi infrastruktur kritis dan data pemerintah. Dalam konteks ini, teknologi Generative AI Tools berbasis Generative Pre-trained Transformer 4 (GPT-4) telah muncul sebagai potensi solusi yang inovatif dalam menghadapi tantangan ini.

GPT-4 multimodal yang dibuat oleh OpenAI, dan merupakan model dasar keempat dalam seri model dasar Generative Pre-trained Transformer (GPT). GPT-4 merupakan model bahasa yang lebih canggih daripada pendahulunya, GPT-3.5, dan memiliki kemampuan untuk menerima input gambar serta teks. GPT-4 dinyatakan lebih andal, kreatif, dan mampu menangani lebih banyak perintah daripada GPT-3.5 (OpenAI, Sam Altman, 2023). GPT-4 juga memiliki kemampuan untuk menghasilkan respon yang lebih aman dan berguna, serta dapat menyelesaikan permasalahan sulit dengan tingkat akurasi yang tinggi. Selain itu, GPT-4 juga dapat menghasilkan skenario serangan yang realistis, yang memungkinkan organisasi untuk mempersiapkan diri dan meningkatkan pertahanan siber mereka.

Generative AI Tool, seperti GPT-4, adalah representasi nyata dari kemajuan pesat dalam kecerdasan artifisial. Kemampuannya untuk memahami bahasa alami, menganalisis teks, dan menghasilkan konten dengan kualitas manusia telah membuatnya menjadi alat yang menjanjikan dalam konteks identifikasi serangan

siber. Meskipun telah banyak penelitian yang mengungkapkan potensinya dalam berbagai aplikasi, penggunaan GPT-4 dalam lingkup pertahanan siber masih memerlukan pemahaman yang lebih mendalam. GPT-4 membuka pintu bagi kemungkinan baru dalam menghadapi serangan siber. Mereka dapat membantu kita mengidentifikasi serangan dengan lebih cepat dan tepat, yang pada gilirannya dapat memperkuat pertahanan nasional kita (Thomas et al, 2023). Dengan pemahaman yang lebih baik tentang potensi dan manfaat Generative AI Tool, diharapkan penelitian ini akan memberikan wawasan yang berharga untuk kemajuan arsitektur keamanan nasional dalam pertahanan siber.

Meningkatkan pertahanan siber guna melindungi dan mempertahankan kedaulatan suatu negara, Indonesia perlu mempertimbangkan pengembangan senjata siber (Hidayati & Gultom, 2019). Serangan siber yang paling umum dan banyak terjadi adalah serangan phishing. Oleh karena itu, pengembangan senjata siber yang memanfaatkan serangan phishing menjadi perhatian dalam upaya mempertahankan kedaulatan suatu negara dengan pengembangan serangan siber melalui phishing. Serangan phishing dengan menggunakan GPT-4 merupakan salah satu bentuk serangan siber yang semakin canggih. Dengan GPT-4 mampu untuk membuat pesan-pesan phishing yang sangat meyakinkan, yang dapat memperdaya target dengan lebih efektif. Sehingga negara dapat mengembangkan senjata siber untuk meningkatkan pertahanan siber guna melindungi dan mempertahankan kedaulatan suatu negara.

Kemajuan teknologi GenAI yang menggunakan GPT-4 untuk serangan phishing menimbulkan tantangan keamanan siber yang signifikan bagi Indonesia, memerlukan langkah-langkah proaktif untuk menciptakan senjata siber yang dapat secara efektif melawan serangan yang semakin kompleks ini dengan memanfaatkan teknologi Generatif AI berbasis GPT-4 yang sama. Oleh karena itu penelitian ‘pemanfaatan Generative AI Tool berbasis GPT-4 dalam melakukan identifikasi serangan simulasi dan skenario dunia nyata untuk membantu meningkatkan keamanan nasional pertahanan siber’ dilakukan. Penelitian ini sebagai alternatif dalam pengembangan senjata siber, dengan tujuan meningkatkan pertahanan siber dan melindungi serta mempertahankan kedaulatan negara. Penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam pengembangan dan keamanan nasional pertahanan siber. Dengan memanfaatkan senjata siber yang difokuskan pada identifikasi dan penanggulangan serangan yang menggunakan Generative AI, seperti GPT-4, dapat

meningkatkan kapasitas pertahanan siber negara. Senjata siber ini diharapkan dapat digunakan untuk melakukan tindakan counter attack yang efektif dalam melawan serangan siber yang semakin kompleks dan berbahaya.

1.2 Identifikasi Masalah

Berdasarkan latar belakang, penulis mengidentifikasi masalah yang menjadi bahan penelitian tesis. Masalah tersebut meliputi:

a. Ancaman siber Generative AI berbasis GPT-4 memiliki resiko seperti spear-phishing, social engineering, deepfakes, plagiarism, privacy concerns, over-reliance, loss of critical thinking, real-time threat detection, the potential for creating new sources karena kurangnya praktik terbaik yang mapan, lemahnya tingkat keamanan, dan hukum. Resiko penggunaan GPT-4 untuk spear-phishing attacks dan social engineering cukup besar, karena bisa menghasilkan konten berpotensi berbahaya, seperti saran untuk merencanakan cyber attacks atau ujaran kebencian.
b. Penulis bertujuan untuk mengeksplorasi efektivitas GPT-4 dalam menghasilkan email phishing persuasif, menggunakan bahasa alat Generatif AI berbasis GPT-4, untuk melihat apakah orang dapat membedakan dan menolak upaya phishing buatan AI atau tertipu oleh email yang dihasilkan AI ini.
c. Alat Generatif AI dalam keamanan siber memerlukan tingkat kemahiran dan ketangkasan yang mendalam, yang dipengaruhi oleh ketersediaan dan pemanfaatan personel terlatih.

1.3 Pembatasan Masalah

Batasan masalah pada penelitian ini adalah:

a. Menggunakan Generatif AI Tool berbasis GPT-4 untuk menciptakan email phishing yang realistis guna untuk memikat dan menarik perhatian banyak orang.
b. Model Perancangan Aplikasi digunakan untuk membantu penulis memetakan penelitian tersebut. Penulis dapat merinci serangan phishing dari awal hingga akhir, termasuk identifikasi serangan dalam simulasi serta analisis serangan dalam skenario dunia nyata setelah email dikirim secara blast phishing dan menerima respon balik.

c. Pada perancangan aplikasi, penulis akan membuat rancangan digunakan untuk mempermudah pembaca apa yang dilakukan penulis dalam penelitian dan eksperimennya. Perancangan ini menggunakan struktur navigasi StarUML dan BPMN Diagram.
d. Dalam penelitian ini menggunakan kerentanan rekayasa social (vulnerability social engineering), user profile, misuse analysis untuk memformulakan prompting.
e. Penelitian ini mengkaji pada User Modeling, Analisis (System Analyst), Hasil Interview.

1.4 Rumusan Masalah

Dengan menggunakan teknologi AI, dapat membantu mengevaluasi sejauh mana pemanfaatan Generative AI Tool's GPT-4 dapat meningkatkan keamanan nasional terhadap serangan phishing email dan apakah teknologi ini dapat menjadi bagian yang signifikan dalam mengoptimalkan arsitektur keamanan pertahanan siber. Rumusan masalah pada penelitian ini dapat dilihat sebagai berikut:

a. Bagaimana efektivitas Generative AI Tool berbasis GPT-4 dalam membangkitkan email phishing yang realistis?
b. Bagaimana formula prompting yang optimal untuk memandu Generative AI berbasis GPT-4 dalam menghasilkan teks email phishing?
c. Bagaimana efektivitas email phising hasil prompting ChatGPT4?

1.5 Tujuan Penelitian

Tujuan penelitian ini adalah untuk mendapatkan pemahaman yang lebih mendalam tentang potensi dan batasan Generative AI Tool berbasis GPT-4 dalam konteks keamanan siber, khususnya dalam pembuatan dan distribusi email phishing. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi penting terhadap pengembangan formulasi kebijakan dan strategi keamanan siber yang lebih efektif dan informasi berharga untuk pencegahan dan peningkatan kesadaran pengguna mengenai risiko serangan phishing.

1.6 Manfaat Penelitian

Manfaat penelitian ini terdiri dari manfaat praktis dan manfaat teoristis:

1.6.1 Manfaat Praktis

1. Riset ini memberikan pemahaman deskriptif yang mendalam mengenai penggunaan Generative AI Tool berbasis GPT-4 dalam mensimulasikan serangan phishing, memungkinkan pemahaman yang lebih baik tentang teknik penipuan canggih ini.
2. Kontribusi utama riset ini terletak pada peningkatan keamanan siber, dimana hasilnya dapat digunakan untuk memperkuat pertahanan siber dan mengoptimalkan alokasi sumber daya.
3. Kemampuan GPT-4 dalam menghasilkan pesan phishing yang meyakinkan, membuka jalan bagi negara-negara untuk mengembangkan strategi pertahanan siber yang lebih efektif dan canggih, membantu dalam perlindungan dan pemeliharaan kedaulatan nasional.
4. Hasil penelitian ini sangat berguna dalam pengembangan program pelatihan keamanan yang lebih efektif, memberikan wawasan kepada personel keamanan untuk mengidentifikasi dan menanggapi serangan phishing yang lebih kompleks.
5. Penelitian ini juga berpotensi dalam penciptaan sistem deteksi otomatis yang lebih efisien, yang dapat mengurangi beban analis keamanan siber dan mempercepat respons terhadap serangan phishing.

1.6.2 Manfaat Teoritis

1. Riset ini dapat menjadi motivator bagi pengembangan teknologi AI yang lebih maju dalam mendeteksi dan melindungi sistem komputer dari serangan phishing.
2. Ini juga memberikan manfaat penting bagi pembuat kebijakan dan pengambil keputusan pertahanan siber, dengan menyediakan informasi mendalam tentang identifikasi serangan dan kesiapan lingkungan dalam menghadapi serangan phishing.
3. Penelitian ini memberikan kontribusi penting ke dalam literatur keamanan siber, dengan memperkenalkan pemahaman baru tentang cara kerja teknologi AI seperti GPT-4 dalam simulasi serangan phishing dan menyajikan konsep teoritis baru dalam studi pertahanan siber.

4. Riset ini berpotensi menjadi sumber referensi yang kaya untuk studi masa depan tentang penggunaan AI Generatif dalam serangan phishing, mendorong inovasi dalam teknik deteksi dan pencegahan serangan.

Dengan demikian, penelitian ini tidak hanya memberikan wawasan praktis yang berharga tetapi juga memajukan pengetahuan dalam bidang keamanan siber, khususnya dalam konteks AI Generatif dan serangan phishing.