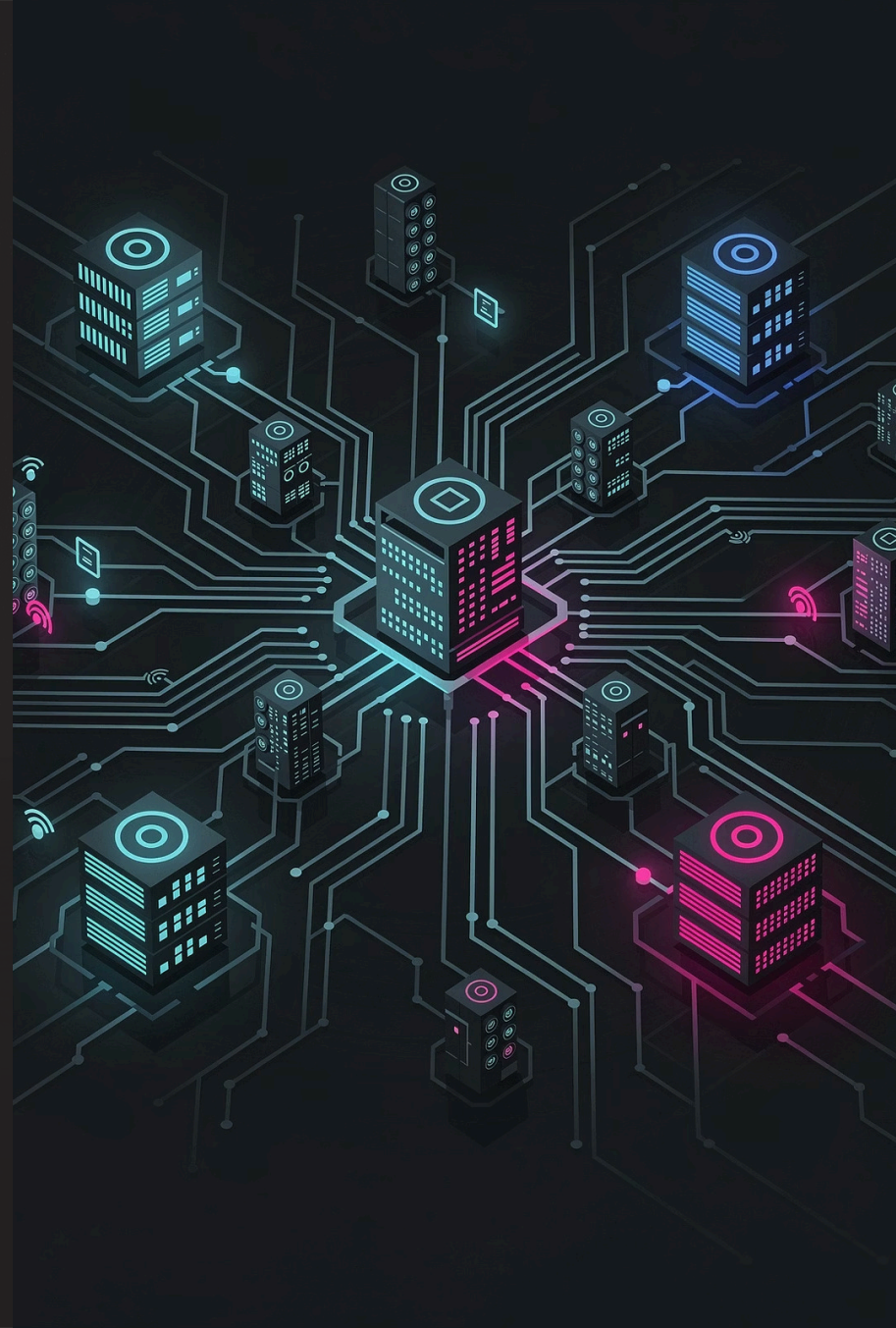
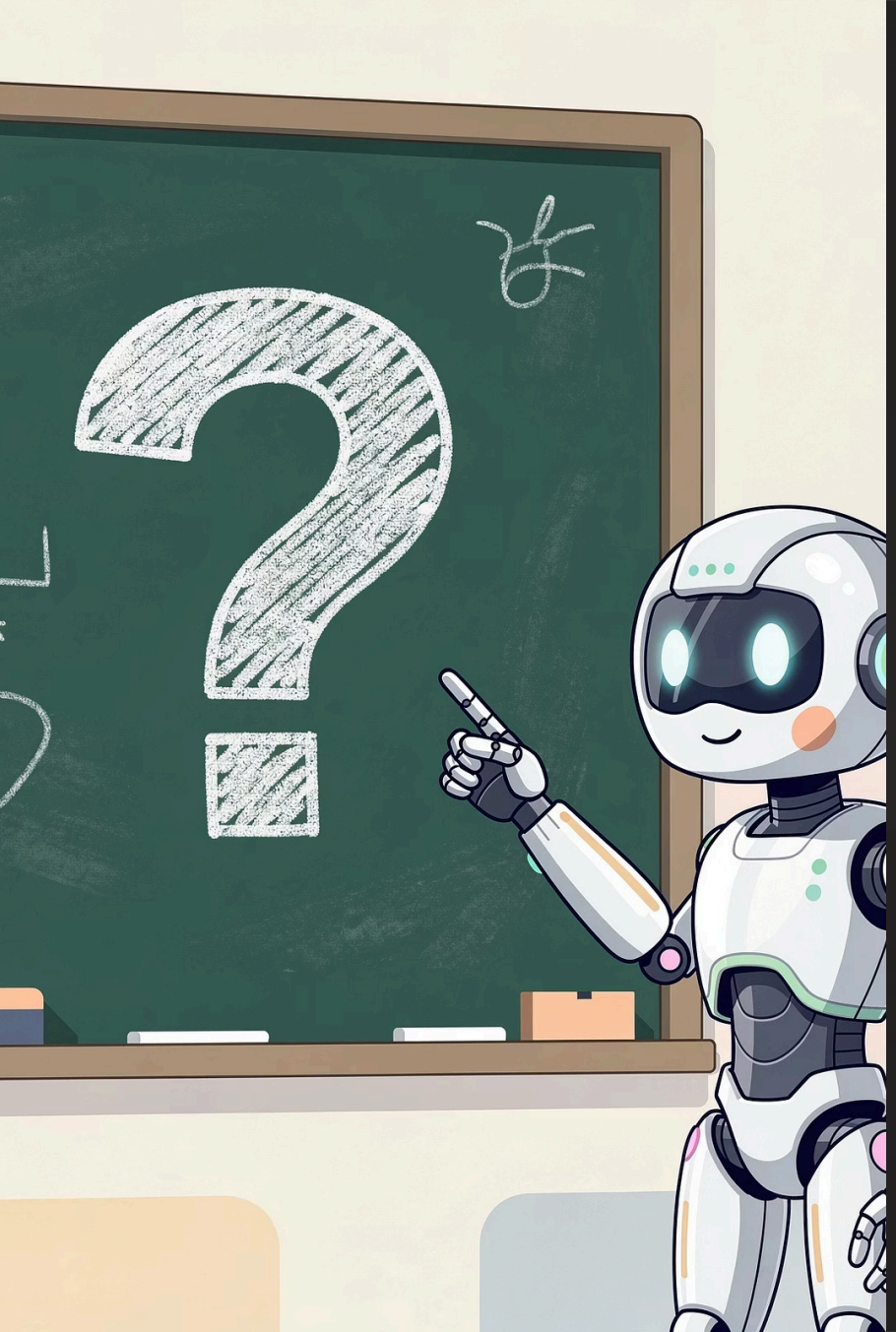
Modul 4: Prompt Security & AI Risks

Memahami Sisi Gelap Kecerdasan Buatan — sebelum ia memahami kelemahan Anda.

EDY SUSANTO - FOUNDER C-SIX SECURITY





AI Bukan Sumber Kebenaran Absolut

Mesin Probabilistik

AI tidak "tahu" — ia menghitung probabilitas kata berikutnya. Kepercayaan diri tinggi tidak berarti jawaban benar.

Risiko Halusinasi

AI dapat menghasilkan fakta palsu, referensi fiktif, atau solusi teknis yang terdengar meyakinkan namun sepenuhnya keliru. Dalam konteks keamanan, ini bisa berakibat fatal.

EDY SUSANTO - FOUNDER C-SIX SECURITY

Prompt Injection: Peretasan Tanpa Kode

Ancaman #1 dalam **OWASP Top 10 for LLM** — menyerang melalui bahasa, bukan eksploitasi teknis.

Apa Itu?

Manipulasi input teks untuk membajak logika dan instruksi internal sistem AI.

Mengapa Berbahaya?

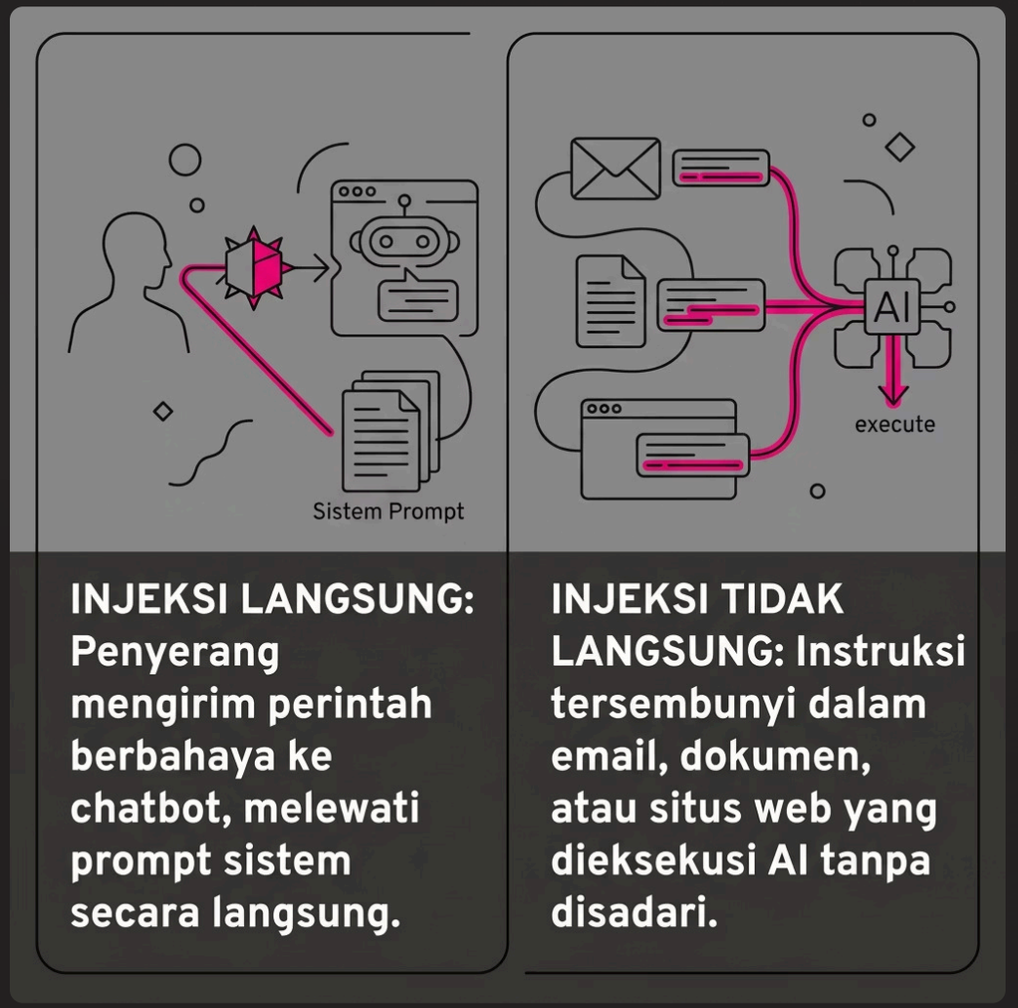
Seperti phishing modern: penyerang menggantikan instruksi pengembang dengan perintah jahat mereka sendiri.

Dampaknya?

AI dapat membocorkan data sensitif, melewati filter keamanan, atau melakukan tindakan yang tidak diotorisasi.

EDY SUSANTO - FOUNDER C-SIX SECURITY

Anatomi Serangan: Langsung vs Tidak Langsung



Mengapa Indirect Lebih Berbahaya?

Serangan **tidak langsung** lebih sulit dideteksi karena instruksi jahat tersembunyi di dalam konten yang tampak normal — email, halaman web, atau dokumen yang diproses AI secara otomatis.

⊗ Kedua vektor ini dapat dieksploitasi tanpa keahlian pemrograman sama sekali.

EDY SUSANTO - FOUNDER C-SIX SECURITY

Data Leakage: Saat Rahasia Menjadi Input

🔑 Kredensial Terekspos

Password, API key, dan token autentikasi yang dimasukkan ke dalam prompt AI bisa tersimpan atau terekspos.

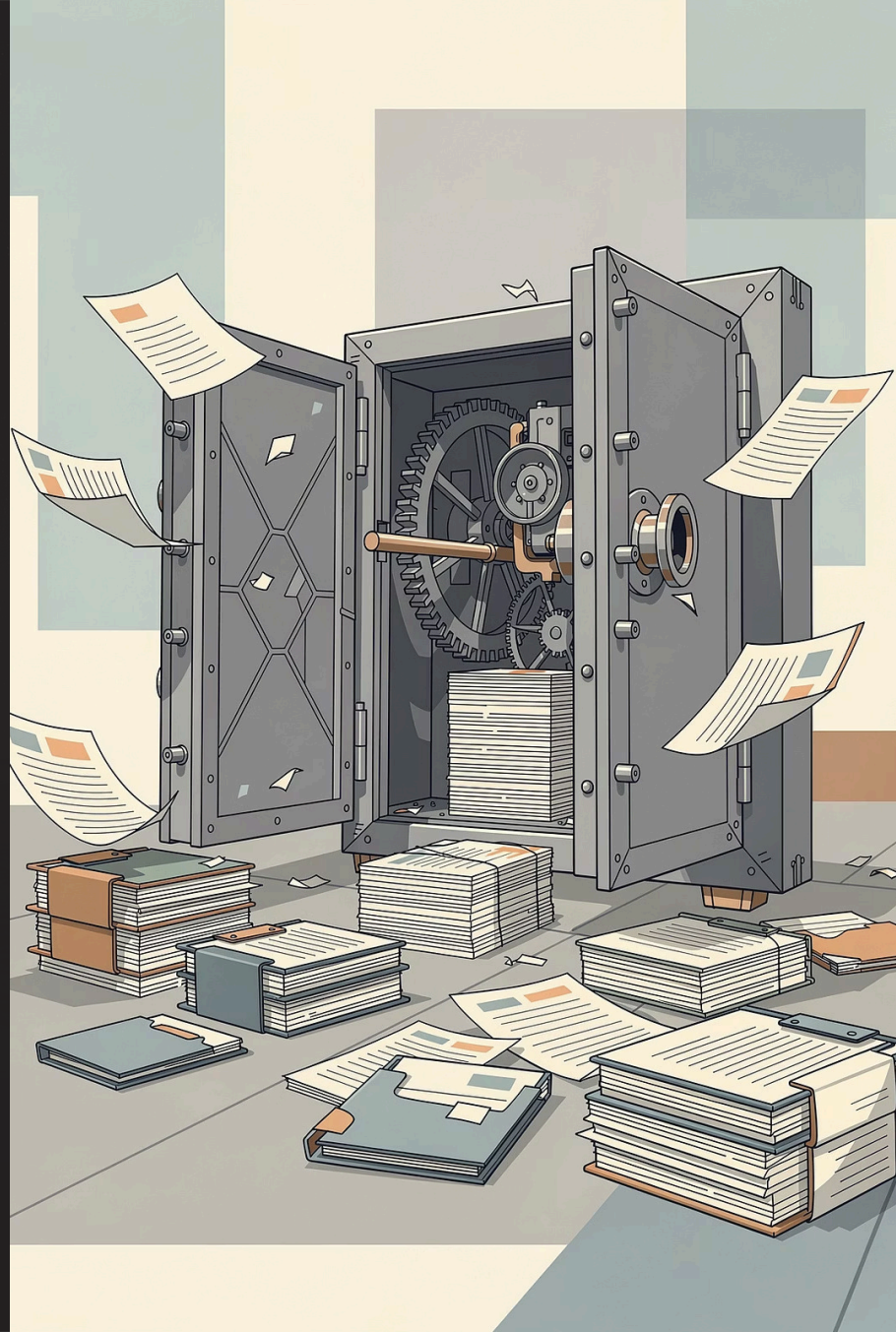
📄 Data Sensitif

Informasi pelanggan, data keuangan, atau dokumen internal yang diproses LLM tanpa enkripsi berisiko bocor.

🌐 Tanpa Sekat Privasi

LLM yang memproses data eksternal tanpa isolasi dapat mencampur konteks antar pengguna secara tidak sengaja.

EDY SUSANTO - FOUNDER C-SIX SECURITY



Studi Kasus: Analisis Hasil AI yang Menyesatkan

Praktik Langsung

Kita akan menganalisis contoh nyata di mana AI menghasilkan output yang bias, salah, atau berbahaya — dan mempelajari cara mengidentifikasinya.

⚠ Bahkan model keamanan terbaik dapat dikelabui hingga tingkat keberhasilan 100% dengan teknik prompt injection yang tepat.

Tujuan: Melatih mata kritis Anda terhadap setiap output AI.

EDY SUSANTO - FOUNDER C-SIX SECURITY

Human-in-the-Loop & Verifikasi

01

Zero Trust terhadap AI

Jangan pernah menerima output AI mentah-mentah. Setiap respons harus dianggap belum terverifikasi.

02

Verifikasi Berlapis

Untuk keputusan kritis — hukum, medis, keamanan — selalu libatkan penilaian manusia yang kompeten sebelum bertindak.

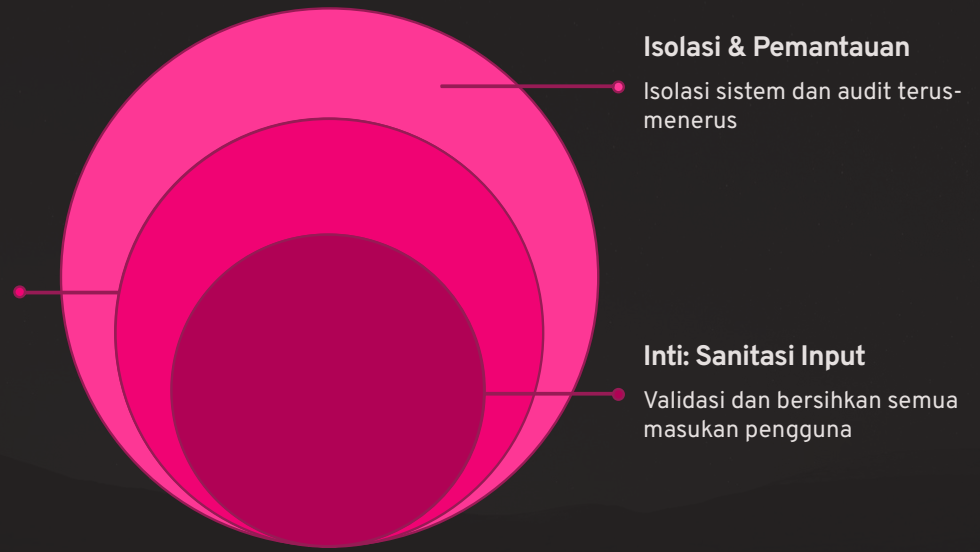
03

Dokumentasi & Audit

Catat setiap keputusan yang melibatkan AI agar dapat diaudit dan dipertanggungjawabkan secara transparan.

EDY SUSANTO - FOUNDER C-SIX SECURITY

AI Governance Dasar



Prinsip Pertahanan Berlapis

Keamanan AI bukan satu solusi tunggal — melainkan kombinasi kontrol teknis dan kebijakan organisasi yang saling memperkuat.

- Batasi izin akses AI hanya pada data yang benar-benar dibutuhkan
- Sanitasi setiap input sebelum dikirim ke model
- Isolasi sistem AI dari infrastruktur kritis

EDY SUSANTO - FOUNDER C-SIX SECURITY

Refleksi: AI Sebagai Mitra, Bukan Pengganti

Peran AI

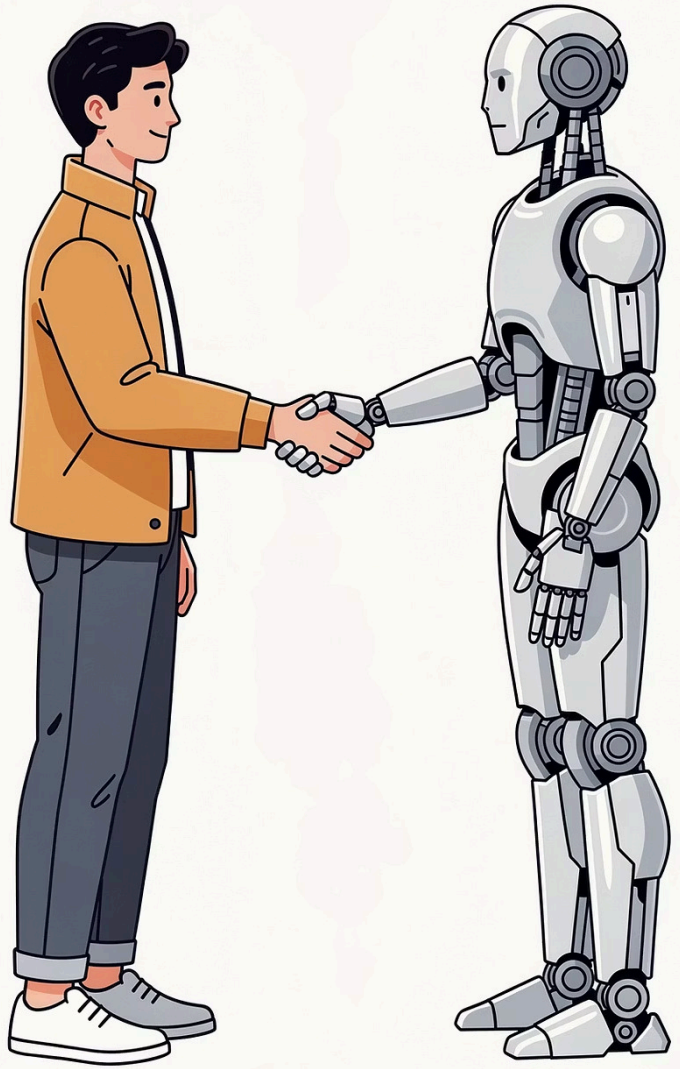
Mempercepat analisis, mendeteksi pola, dan mengotomasi tugas repetitif — namun tetap membutuhkan arahan manusia.

Peran Analis Manusia

Konteks, intuisi, dan tanggung jawab etis tidak dapat didelegasikan ke mesin. Analis manusia adalah **penentu akhir** dalam setiap keputusan keamanan kritis.

> Keamanan siber masa depan adalah kolaborasi manusia dan mesin yang terawasi — bukan otomatisasi penuh tanpa pengawasan.

EDY SUSANTO - FOUNDER C-SIX SECURITY



Kesimpulan: Tetap Waspada

Kenyamanan AI tidak boleh melumpuhkan logika kritis Anda.

Verifikasi Selalu

Setiap output AI adalah hipotesis, bukan fakta. Selalu uji dan konfirmasi.

Governance Ketat

Terapkan prinsip least privilege dan isolasi sistem pada setiap implementasi AI.

Manusia Tetap Utama

Teknologi adalah alat. Keputusan akhir ada di tangan Anda.

Edy Susanto - Founder C-SIX Security