BO·TT·EGA
IT minds

Program szkolenia:

AI Security and Compliance: Ochrona Danych, Prawo i Cyberzagrożenia

Informacje:

Nazwa:	**AI Security and Compliance: Ochrona Danych, Prawo i Cyberzagrożenia**
Kod:	**ai-security**
Kategoria:	AI
Czas trwania:	1 dzień
Forma:	50% wykłady / 50% warsztaty

Wdrożenie AI w firmie to nie tylko szansa na zysk, to także otwarcie nowych wektorów ataku.

To szkolenie pokrywa dwa krytyczne obszary:

- Legal and Governance: Jak legalnie i bezpiecznie używać AI, aby nie naruszyć RODO, nie stracić tajemnicy przedsiębiorstwa i zachować kontrolę nad danymi.
- AI-Driven Threats: Jak bronić się przed atakami nowej generacji, gdzie hakerzy wykorzystują AI do tworzenia hyper-realistycznego phishingu, deepfake'ów głosowych i wideo oraz zatruwania danych.

Celem nie jest blokowanie innowacji, ale budowa świadomości. Uczymy, jak odróżnić narzędzia bezpieczne (Enterprise) od publicznych i jak nie stać się ofiarą "AI social engineeringu".

Szkolenie łączy perspektywy:

- Prawników i Compliance: Aspekty regulacyjne i odpowiedzialność
- Pracowników Biurowych i Grafików: Bezpieczne narzędzia i weryfikacja treści (Deepfakes)
- IT i Security: Nowe wektory ataków (Malware delivery, Poisoning).

Zalety szkolenia:

- Compliance: Zasad klasyfikacji danych – co wolno wrzucić do ChatGPT, a co jest surowo zabronione.
- Higieny Cyfrowej: Różnic między modelami publicznymi a prywatnymi instancjami (Enterprise) w kontekście RODO i Audit Trail.
- Wykrywania Ataków: Rozpoznawania phishingu tworzonego przez AI (bez błędów językowych, z kontekstem) oraz Deepfake'ów (audio/wideo)
- Human-in-the-Loop: Dlaczego człowiek musi pozostać ostatecznym decydentem i jak wygląda odpowiedzialność prawna za błędy AI.
- Zagrożeń dla Danych: Czym jest Data Poisoning i jak atakujący mogą manipulować wynikami modeli

BO·TT·EGA
IT minds

Szczegółowy program:

1. AI Governance i Aspekty Prawne (Fundamenty)

1.1. Public vs. Enterprise: Dlaczego darmowy ChatGPT to koszmar dla działu bezpieczeństwa? Analiza regulaminów, retencji danych i trenowania modeli na danych użytkownika.

1.2. Klasyfikacja Danych: Warsztat praktyczny – "Traffic Light Protocol". Ustalenie zasad, jakie typy dokumentów (dane osobowe, finansowe, kod źródłowy) mogą być przetwarzane przez AI.

1.3. RODO i AI Act: Automatyzacja a prawa jednostki. Czym jest Audit Trail (ślad rewizyjny) i dlaczego zasada Human-in-the-Loop jest niezbędna do uniknięcia odpowiedzialności karnej/cywilnej.

2. Shadow AI i Wycieki Danych

2.1. Ryzyko "Shadow AI": Pracownicy używający nieautoryzowanych narzędzi AI do skracania sobie pracy. Jak to wykrywać i jak temu zapobiegać (edukacja vs. blokady).

2.2. Prompt Injection: Jak nieświadomy pracownik może zostać zmanipulowany przez model lub jak model może zdradzić instrukcje systemowe.

2.3. Case Study: Analiza głośnych wycieków danych spowodowanych przez wklejenie kodu lub strategii do publicznych chatbotów.

3. Ofensywa AI – Phishing i Social Engineering 2.0

3.1. AI-Driven Phishing: Koniec z "listami od nigeryjskiego księcia". Jak AI tworzy perfekcyjne językowo, spersonalizowane wiadomości (Spear-phishing) na podstawie danych z LinkedIn/social media.

3.2. Generowanie złośliwych treści: Evil URLs i wsparcie AI w pisaniu kodu malware'u. Jak atakujący omijają filtry bezpieczeństwa.

3.3. Obrona: Na co zwracać uwagę w erze, gdy styl i gramatyka nie są już wyznacznikiem oszustwa? Weryfikacja kontekstowa.

4. Deepfakes i Dezinformacja (Audio/Video)

4.1. Klonowanie Głosu (Vishing): Jak 3 sekundy nagrania wystarczą, by podszyć się pod prezesa (CEO Fraud). Procedury weryfikacji tożsamości w rozmowach telefonicznych.

4.2. Deepfake Wideo: Zagrożenia dla wizerunku marki i manipulacje na spotkaniach online.

4.3. Dla grafików i marketingu: Odpowiedzialność za oznaczanie treści generowanych przez AI (watermarking) i ryzyka prawne związane z prawami autorskimi.

5. Ataki na Infrastrukturę Danych (Techniczne)

5.1. Data Poisoning: Jakie są skutki celowego "zatrucia" danych treningowych? Sabotaż procesów decyzyjnych w firmie.

5.2. Model Poisoning and Adversarial Attacks: Jak zmylić systemy rozpoznawania obrazu lub klasyfikacji tekstu (np. omijanie filtrów spamu, oszukiwanie systemów wizyjnych).

5.3. Podsumowanie: Budowa procedur bezpieczeństwa i kultury Zero Trust wobec treści cyfrowych