AI Kybernetické útoky
&
AI Bezpečnost aplikací

Vilém Markovič

comma_0

comma_0

AI-POWERED

Kybernetické útoky

Jak umělá inteligence mění krajinu hrozeb

+442 %

nárůst AI vishingu (2025)

25 M USD

ztráta Arup deepfake (2024)

30-100 mld USD

globální škody 2025

Přehled druhů AI-poháněných útoků

Klasifikace dle techniky a cíle

Deepfake Video/Hlas

GAN, voice cloning – CEO fraud, převody peněz

AI Phishing/Spear

LLM generování, 54% CTR vs 12% klasický

Polymorfní Malware

AI mutace kódu, obejití AV signatur

AI Zero-Day Discovery

Automatizovaný fuzzing, exploit v minutách

Data Poisoning

Manipulace trénovacích dat AI modelů

Ransomware + AI Recon

ML průzkum, AI negotiation, double extortion

AI snižuje bariéry pro útočníky: čas přípravy útoku se zkrátil z měsíců na hodiny

Deepfake & Voice Cloning – Reálné případy

Ztráty v řádu desítek milionů USD

2024 **ARUP (Hong Kong)**

Ztráta: 25 M USD

Deepfake video konference

Zaměstnanec převedl 25 mil. USD po videohovoru, kde všichni kolegové včetně CFO byli AI deepfaky. Útočníci nahrávku připravili z veřejně dostupných videí.

2024 **WPP / CEO Mark Read**

Ztráta: Odhaleno

AI voice clone – Teams hovor

Klonován hlas CEO Marka Reada v Microsoft Teams. Útok odhalen díky osobní kontrolní otázce. Použity veřejné nahrávky z konferencí a podcastů.

2024 **Ferrari / CEO Vigna**

Ztráta: Odhaleno

Voice clone na WhatsApp

Naklonován hlas CEO Benedetta Vigna s napodobením jeho přízvuku na WhatsAppu. Odhalen otázkou ohledně specifické smlouvy, kterou útočník neznal.

2021 **Ozy Media**

Ztráta: Pokus 40 M USD

Voice clone – Goldman Sachs

Manažer naklonoval hlas YouTube partnera při investičním hovoru s Goldman Sachs. Jeden z prvních zdokumentovaných případů AI voice clone ve finanční kriminalitě.

AI-generovaný Malware & Ransomware

TA547, Medusa, Play – reálné kampaně 2024–2025

TA547 / Rhadamanthys

DE firmy · 2024

– Polymorfní AI loader – každé spuštění unikátní kód

– Doručuje Rhadamanthys infostealer (hesla, tokeny)

– Německé organizace – cílené spear-phishing kampaně

– AI zkrátila vývoj nové varianty z týdnů na hodiny

Play Ransomware + AI Zero-Day

Kritická infrastruktura · 2025

– CVE-2025-29824 – AI-nalezený zero-day v Windows driveru

– Eskalace privilegií, deaktivace antiviru

– Šifrování VMware ESXi virtuálních strojů

– Double-extortion přes Tor + data leak stránka

Medusa Ransomware RaaS

300+ obětí USA · 2023–2025

– AI recon – automatický průzkum CVE v prostředí oběti

– MOVEit exploit (CVE-2023-34362) + další 0-day

– Exfiltrace dat před šifrováním (double extortion)

– Cílí kritická infrastruktura: health, vzdělání, finance

Storm-0817 / ChatGPT Android Spy

Íránský aktér · 2024

– Íránský APT použil ChatGPT k vývoji Android spywaru

– Krádež SMS, kontaktů, lokace, souborů a hesel

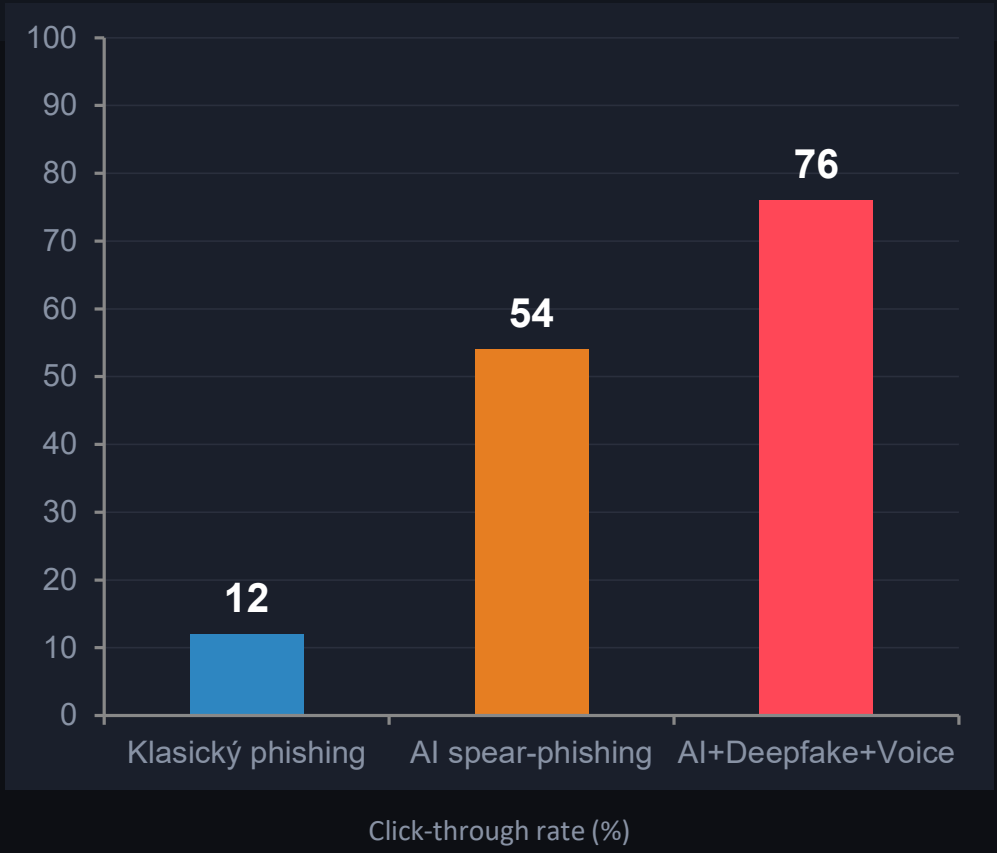
– AI zkrátila dobu vývoje z měsíců na dny

– Platforma ChatGPT zneužita a přístup zablokován

Společný jmenovatel: AI výrazně zrychluje vývoj, mutaci a deployment malware – detekce signaturami selhává

AI Phishing – Statistiky a technika

LLM-generované zprávy vs. klasický phishing



Personalizace v reálném čase

LLM generuje zprávu přizpůsobenou konkrétní osobě – jméno, role, kontext z LinkedIn / OSINT

Bezchybná čeština / angličtina

AI odstraňuje typické jazykové chyby phishingu – zprávy jsou nerozpoznatelné od legitimní komunikace

320 firem za rok – 1 kampaň

CrowdStrike zdokumentoval kampaň využívající GenAI v celém kill-chainu: recon → phish → exploit → persist

Průměrná cena AI-poháněného breache: 5,72 mil. USD (+13 % YoY) · Globální škody: 30 mld USD ročně

Jak se bránit AI útokům

Vrstvená obrana – technická + procesní

Anti-Deepfake

- Ověřovací protokol pro finanční transakce (code-word)
- Real-time deepfake detekce (Microsoft DART, Sensity)
- Školení zaměstnanců – neobvyklé požadavky vždy ověřit

AI-Aware Email Security

- AI-powered email filtering (Proofpoint, Mimecast AI)
- Behavioral analytics – detekce anomálního odesílatele
- DMARC/DKIM/SPF enforcement – zamezení spoofingu

Ochrana endpointů (EDR)

- Behavioral AI – detekce bez signatur (Cortex XDR, CS)
- Polymorfní malware neobejde behaviorální analýzu
- Karanténa + forenzní snapshot při detekci

Threat Intelligence

- Unit 42 / CrowdStrike Intel – 150M+ IoC denně
- MITRE ATT&CK mapování každého alertu
- XDR korelace cross-layer – propojení izolovaných eventů

Identity & Zero Trust

- MFA povinně pro všechny – phishing-resistant (FIDO2)
- Conditional Access + device compliance enforcement
- Privileged Access Management (PAM) pro kritické systémy

AI Security Governance

- Politika použití AI nástrojů (schválené vs. stínové AI)
- Data Loss Prevention pro AI výstupy
- Pravidelné red-team cvičení simulující AI útoky

comma_0

NOVINKA

24. března 2026 mezi 10:39–16:00 UTC

LITE LLM 1.82.7 a 1.82.8

Kompromitované verze Python

LiteLLM AI Gateway vyšetřuje podezření na supply chain útok, který zahrnoval neoprávněné publikování balíčků na PyPI.
Současné důkazy naznačují, že mohl být kompromitován účet správce na PyPI a zneužit k distribuci škodlivého kódu.
V tuto chvíli se domníváme, že incident může souviset s širším bezpečnostním problémem kolem nástroje Trivy, kde byly odcizené přihlašovací údaje použity k neoprávněnému přístupu do publikačního pipeline LiteLLM.
Vyšetřování stále probíhá a uvedené informace se mohou změnit.

comma_0

BEZPEČNÉ NASAZENÍ

AI Aplikací

Požadavky EU AI Act a bezpečnostní standardy

Pro dodavatelské firmy vyvíjející AI aplikace v cloudu

16. 2. 2026

EU AI Act – Základy a klasifikace rizik

Nařízení (EU) 2024/1689 – platné od 2025

NEPŘIJATELNÉ RIZIKO

ZAKÁZÁNO

VYSOKÉ RIZIKO

PŘÍSNÁ PRAVIDLA – finanční sektor, GDPR zvláštní kategorie

NÍZKÉ / MINIMÁLNÍ RIZIKO

TRANSPARENTNOST + základní opatření – veřejná data

Finanční sektor = VYSOKÉ RIZIKO:

klienti a úvěruschopnost · schvalování půjček · profilování rizik zákazníků · provoz kritické finanční infrastruktury

⚠ Klasifikace závisí na USE CASE, ne na technologii!

Reálná situace – 2 typy prostředí

Připravenost na bezpečný deployment AI aplikací

✓ Striktně řízená bezpečnost & SDLC

- Prostředí striktně ohraničeno, hardening zajištěn
- Git – jasná governance a schvalovací proces
- Artifactory – komplexní bezpečnostní scan artefaktů
- Zdroj dat – dedikovaná data pro schválený use case
- Azure – striktní Policy, APIM, AI + Blob storage řízeny
- Zákazník nepustí mimo mantinely → Nízké riziko

✗ Vše na nízké úrovni – a přes to potřeba AI

- ✗ Bez bezpečnostní politiky pro AI
- ✗ Bez bezpečnostního dohledu a auditování
- ✗ Bez security review procesu
- ✗ Bez klasifikace dat a přístupových kontrol
- ✗ Bez incident response plánu
- ✗ Vysoké riziko porušení EU AI Act a GDPR

9 Hlavních bezpečnostních hrozeb AI aplikací

OWASP LLM Top 10 – adaptace pro finanční prostředí

1 **Prompt Injection**

Manipulace modelu škodlivými vstupy → úniky dat

2 **Otrávení dat**

Manipulace trénovacími daty → zkreslené výstupy

3 **Odhalení citlivých dat**

Únik PII, GDPR dat z kontextového okna

4 **Krádež modelu**

Reverse-engineering, distillation útoky

5 **Selhání modelu**

Halucinace, bugs, nepředvídatelné chování

6 **Supply chain**

Kompromitované knihovny a závislosti

7 **Nesecure pluginy**

Zranitelné extensions a integrace

8 **Denial of Service**

Přetížení, nadměrné cloudové náklady

9 **Ztráta compliance**

Porušení GDPR, EU AI Act, ISO 27001

Povinná opatření – Vysoké riziko & API Security

OAuth 2.0 · mTLS · Anti-injection · Model Integrity

ID	Požadavek	Detail
1.10.1	Uchovávání dat min. 6 měsíců	Secure environment, access controls
2.1	API zabezpečení	OAuth 2.0, mTLS, HTTPS, rate limiting
2.2	Validace vstupů/výstupů	Anti-prompt injection, sanitizace
1.1	Kryptografické ověření	Digital signatures, checksums, integrity
1.2	Archivace modelů	Encrypted copies, tamper-proof storage
1.6	Supply chain security	Vendor assessment, SOC 2, model verification

Autentizace

OAuth 2.0 + JWT tokens · Mutual TLS (mTLS) · API keys s rotací · Service principals · Managed Identity (Azure)

Autorizace

RBAC – Role-Based Access · Principle of Least Privilege · Rate limiting per user/IP · Azure API Management · Continuous monitoring

Pre-Deployment Checklist

11 oblastí pro bezpečné nasazení AI aplikace

Compliance
GDPR, ISO 27001 ověřeno

Klasifikace
Systém klasifikován dle EU AI Act

Dokumentace
Risk assessment zdokumentován

API Security
OAuth 2.0 + mTLS implementováno

Input Validation
Anti-injection filtry aktivní

Model Integrity
Hash verification + signatures

Data Retention
Min. 6 měsíců pro high-risk

Monitoring
Azure Monitor + alerty nastaveny

Encryption
At rest + in transit

Access Control
RBAC + Least Privilege

Supply Chain
Vendor assessment dokončen

Azure Best Practices pro AI

Bezpečnostní konfigurace klíčových Azure služeb

Azure OpenAI

Private Endpoints · Content Filtering · Managed Identity · No public access

Azure Blob Storage

CMK encryption · Versioning · Soft delete · Private Link · Immutable storage

Azure Key Vault

Purge protection · RBAC · Soft delete 90d · HSM · Access logging

Azure API Management

Rate limiting · OAuth validation · IP filtering · DDoS protection

Azure Policy

Compliance enforcement · Drift detection · Auto-remediation · Deny assignments

Defender for Cloud / CNAPP

CSPM · Vulnerability scanning · Threat protection · Security score

Azure Monitoring

Custom alerts · Log Analytics · Action Groups · Dashboards · Retention policy

Azure AD / Entra

Conditional Access · MFA mandatory · Device compliance · PIM for privileged roles

NEDĚLAT – Časté chyby a jejich dopad

8 nejčastějších selhání při nasazení AI aplikací

✗ Hardcoded API keys v kódu

→ Únik credentials, security breach

✗ Veřejné API endpointy bez autentizace

→ Zneužití, nadměrné náklady

✗ Bez validace vstupů

→ Prompt injection, manipulace modelu

✗ Model bez integrity checks

→ Neodetekovaná modifikace

✗ Nedostatečné logování

→ Nemožnost forensics a auditu

✗ Přímý přístup k production

→ Absence staging, rizikové změny

✗ Vendor bez due diligence (bezpečnostní ověření)

→ Supply chain attack

✗ Ignorování EU AI Act

→ Pokuty až 7 % globálního obratu

Užitečné zdroje & kontakt

EU AI Act

https://www.ey.com/en_ch/insights/forensic-integrity-services/the-eu-ai-act

ENISA Cybersecurity AI

https://www.enisa.europa.eu/publications/multilayer-framework

Azure AI Security

https://learn.microsoft.com/azure/ai-services/security

NIST AI Framework

https://airc.nist.gov/AI_RMF_Knowledge_Base/AI_RMF

CSA AI Security

https://cloudsecurityalliance.org/research/artifacts?term=ai

Databricks AI Security

https://www.databricks.com/resources/whitepaper/dasf

❓ **Otázky a odpovědi · Děkujeme za pozornost!**

✉ Kontakt vilem.markovic@comma0.io