AI a kyberkriminalita

Ing. Lubomír Záliger

Samostatné oddělení kybernetické bezpečnosti

Ministerstvo vnitra

9. dubna 2025

Klasifikace: **Veřejné**, **TLP:CLEAR**

Kybernetická bezpečnost

- Ministerstvo vnitra má v kybernetické bezpečnosti České republiky důležitou a nezastupitelnou roli a spolupracuje na strategii kybernetické bezpečnosti s Národním úřadem pro kybernetickou a informační bezpečnost.
- Primárním úkolem Ministerstva vnitra je ochrana informačních systémů státu a boj s kybernetickou kriminalitou.
- Vývoj umělé inteligence a strojového učení je v současnosti velmi dynamický proces a samostatné oddělení kybernetické bezpečnosti monitoruje aktuální trendy vývoje.
- AI má v oblasti kybernetické bezpečnosti dvě zásadní role – může být využita k prevenci, ochraně a obraně, ale také zneužita k útokům.

Typickými případy kybernetické kriminality jsou:

Počítačové podvody – phishing, krádež identity, podvody s kreditními kartami, krádež financí.

Malware – škodlivý software, viry, trojské koně, ransomware, spyware.

Hacking – neoprávněný přístup k počítačovým systémům.

Neoprávněné stahování a sdílení dat – porušování autorských práv k filmům, hudbě, softwaru.

Kybernetická kriminalita v oblasti osobní bezpečnosti a sociálních rizik je:

Kyberšikana – forma šikany, která se odehrává v on-line prostředí.

Stalking a kyberstalking – sledování a obtěžování jednotlivců.

Sexting – posílání nebo přijímání sexuálně explicitních zpráv nebo obrázků.

Kybergrooming – dospělí navazují kontakt s dětmi on-line s cílem sexuálního zneužití.

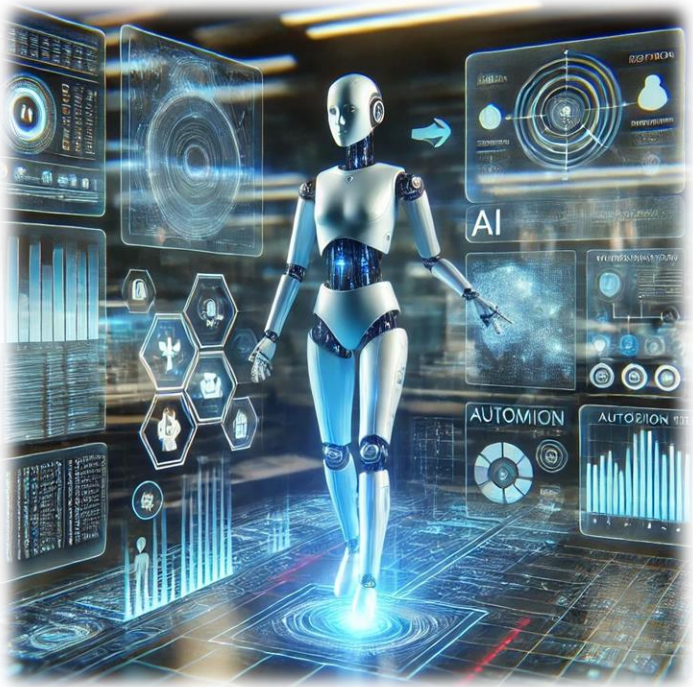
Závažnou kybernetickou kriminalitu představuje:

Kybernetická špionáž – získávání citlivých informací, cílena na vládní instituce, korporace nebo i jednotlivce.

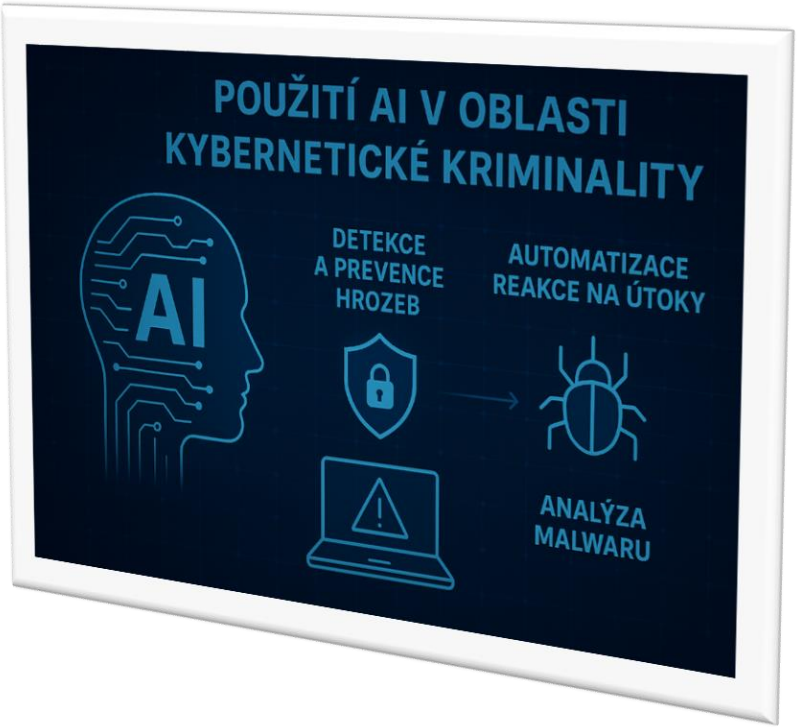
Kyberterorismus – kybernetické útoky, zastrašování nebo poškozování společností, vlád nebo jednotlivců.

Aplikace AI v oblasti kybernetické bezpečnosti

- Integrace AI do bezpečnostních systémů
- Ochrana před pokročilými hrozbami
- Automatizace reakce na hrozby
- Prediktivní analýza
- Pokročilé strojové učení
- Deep Learning a neuronové sítě
- Kvantové výpočty



- Detekce a prevence hrozeb
- Automatizace reakce na útoky
- Analýza malwaru
- Behaviorální analýza
- Prediktivní analýza
- Phishing a sociální inženýrství
- Deepfake podvody
- Dezinformace (hybridní hrozby)



Etické otázky spojené s tréninkem AI

- Transparentnost a vysvětlitelnost (vysvětlení proč a jak AI dospěla k nějakému závěru).
- Předpojatost a diskriminace (předsudky obsažené v tréninkových datech).
- Ochrana soukromí (narušení soukromí až krádež identity).
- Zodpovědnost (kdo ponese odpovědnost za chybné rozhodnutí – etická a právní rovina).
- Autonomie (nekontrolovatelný následek např. ve vojenském sektoru).
- Pracovní trh a ekonomické dopady (ztráta pracovního místa – kompenzace státem, sociální bouře, celospolečenský dopad).

Zneužití:

- škodlivé aktivity (kybernetické útoky),
- šíření dezinformací,
- vytváření deepfake videí,
- narušení integrity informací,
- adaptivní malware,
- autorská práva.

Ochrana soukromí:

- možnost shromažďování a analyzování obrovského množství osobních dat,
- ochrana soukromí - obavy o zneužití osobních dat.

Sociální pohled na kybernetickou kriminalitu

- Kybernetická kriminalita představuje vážnou hrozbu v sociální oblasti nejen pro jednotlivce, ale i pro celou společnost.
- **Ztráta důvěry:** kybernetická kriminalita může vést k celkové ztrátě důvěry a poklesu využívání digitálních technologií a on-line služeb.
- **Zvýšená regulace:** rostoucí počet kybernetických útoků vede k přísnějším regulacím a zákonům, které mají za cíl zvýšit bezpečnost a ochranu dat.

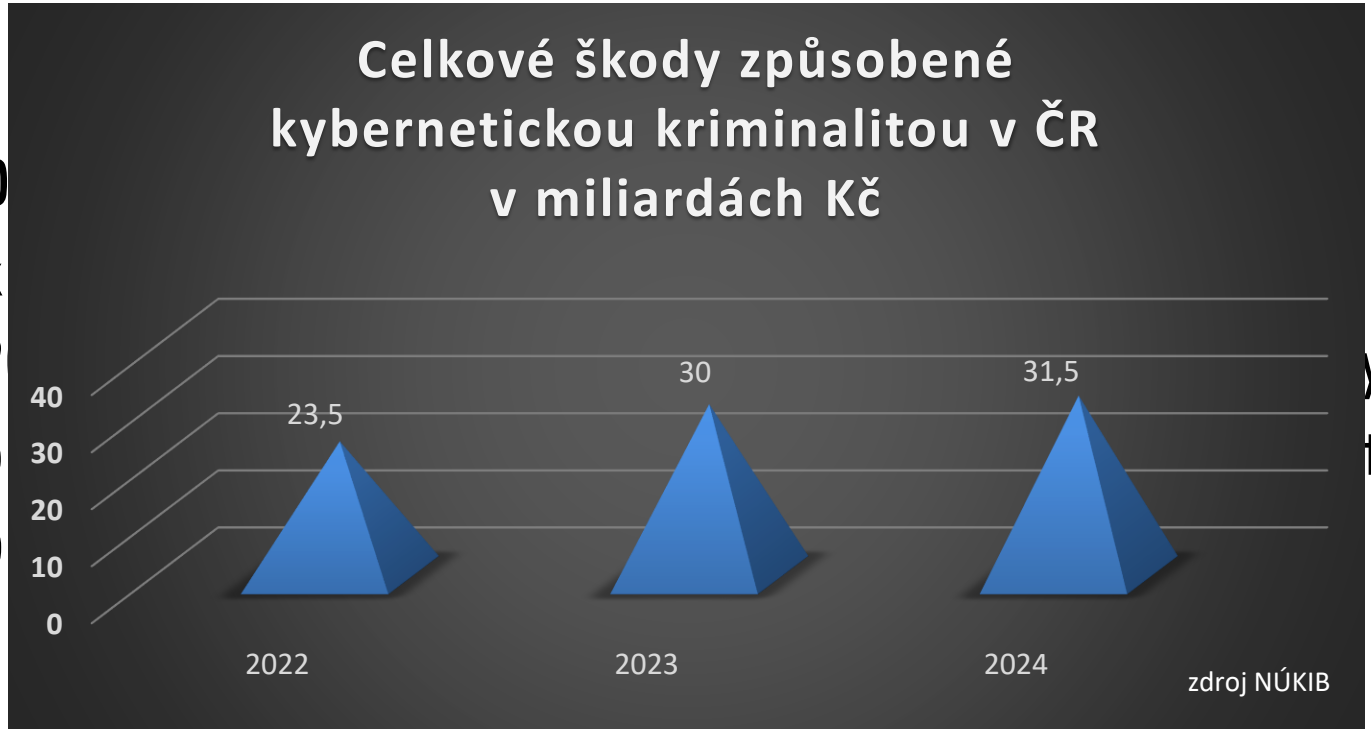
Kybernetická kriminalita má výrazné ekonomické dopady na jednotlivce:

1) **finanční ztráty:** jednotlivci přijdou o peníze v důsledku podvodů, krádeže identity nebo neoprávněných transakcí (podvody s kreditními kartami), často se dostávají do dluhů,
2) **ztráta osobních údajů:** kybernetické útoky mohou vést k úniku citlivých osobních údajů, což může mít dlouhodobé důsledky pro soukromí a bezpečnost jednotlivců,
3) **psychologické dopady:** oběti kybernetické kriminality často zažívají stres, úzkost a pocit bezmoci, strach z dalšího útoku může negativně ovlivnit jejich každodenní život.

Kybernetická kriminalita má významné ekonomické dopady na firmy:

1) **výpadky služeb:** kybernetické útoky mohou způsobit výpadky kritických služeb, což může mít vážné důsledky především pro veřejný sektor, zdravotnictví a další klíčové oblasti,
2) **poškození reputace:** firmy, které se stanou obětí kybernetických útoků, mohou utrpět poškození své pověsti, ztráta důvěry zákazníků může vést k poklesu příjmů včetně dlouhodobých negativních dopadů na obchodní vztahy,
3) **investice do bezpečnosti:** firmy musí investovat značné prostředky do kybernetické bezpečnosti, aby se chránily před útoky (náklady na nové technologie, školení zaměstnanců, pravidelné bezpečnostní audity),
4) **regulační sankce:** porušení předpisů (např. v oblasti GDPR), může vést k vysokým pokutám a zvýšení pojistných nákladů kvůli vyššímu riziku.

- O ckou
  k roku
  2 y na
  o tické
  b



AI a budoucnost

- **Optimistický scénář:** AI řeší klimatické krize, nemoci, chudobu, lidé žijí v éře bez práce s univerzálním příjmem.
- **Pesimistický scénář:** AI způsobí kolaps společnosti (události typu „Terminátor“ nebo „Matrix“).
- **Smíšený scénář:** AI kombinuje přínosy i rizika, AI se stane mocným nástrojem, ale lidská kontrola a etika zůstanou klíčové, úspěch bude závislý na rovnováze mezi technologickým pokrokem a odpovědným řízením.

Budoucí vývoj – trendy

- Turbulentní vývoj ve směru růstu a rozvoje AI (soupeření USA, Čína)
- Využití AI v oblasti vzdělávání a práce (personalizované učení, automatizace profesí jako administrativa, call centra)
- Specializovaná AI (výzkum, farmacie, lékařství)
- Běžné využití AI v menších firmách i jednotlivci
- AI bez on-line připojení na internet
- Zdokonalený malware a škodlivý kód obecně
- Zdokonalený phishing a deepfake
- Zdokonalená e-komerce (cílené reklamy)

AI a budoucnost

- Turingův test překonán bez různých „ale“.
- GPT 4.5 dokáže hrát člověka hyper-realisticky, realističtějšího než jsou skuteční lidé.

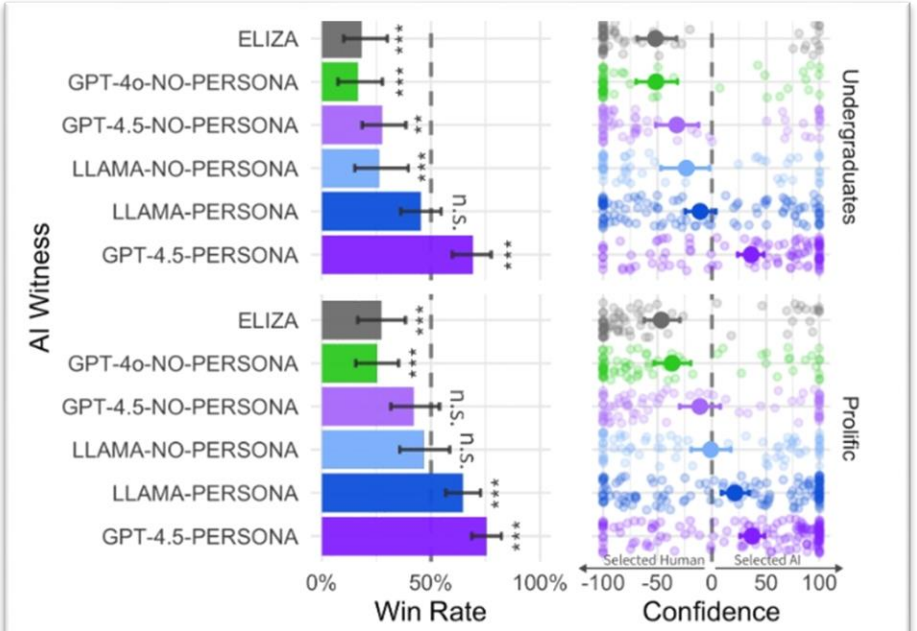


Figure 2: **Left:** Win rates for each AI witness: the proportion of the time that the interrogator judged the AI system to be human rather than the actual human witness. Error bars represent 95% bootstrap confidence intervals. Asterisks next to each bar indicate whether the win rate was significantly different from chance (50%). **Right:** Confidence in verdicts where the interrogator selected the actual human or the AI model for each witness type. Each point represents a single game. Points further toward the left and right indicate higher confidence that the AI is the AI versus the human respectively. Error bars indicate 95% bootstrap confidence intervals around the mean.

zdroj: http://arxiv.org/pdf/2503.23674

Děkujeme za pozornost.