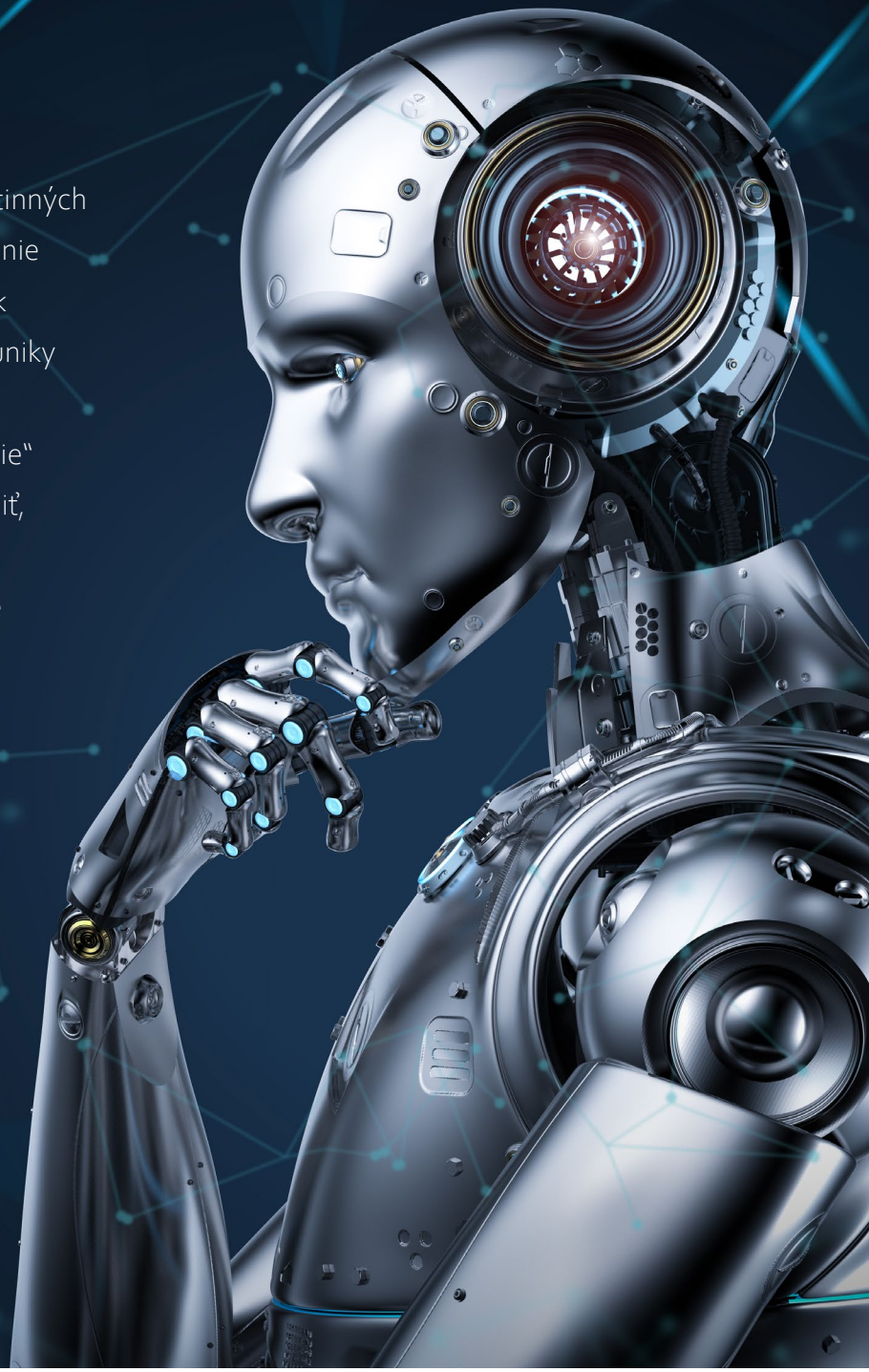
UMELÁ INTELIGENCIA

NIE JE POSTRACHOM PRE KYBERNETICKÚ BEZPEČNOSŤ

Umelá inteligencia (AI) prináša firmám významné benefity — automatizáciu rutinných úloh, podporu rozhodovania, zefektívnenie procesov a tvorbu obsahu. Súčasne však predstavuje nové bezpečnostné riziká: úniky údajov, zneužitie modelov, generovanie škodlivého kódu, deepfake či „halucinácie" modelov. Cieľom tohto článku je vysvetliť, čo AI je a čo nie je, prečo ju firmy musia používať, aké riziká so sebou nesie a aké konkrétne opatrenia a pravidlá by mali organizácie zaviesť, aby minimalizovali bezpečnostné hrozby.



Spolufinancovaný Európskou úniou

Financované Európskou úniou. Vyjadrené názory a postoje sú názormi a vyhláseniami autora(-ov) a nemusia nevyhnutne odrážať názory a stanoviská Európskej únie. Európska únia za ne nepreberá žiadnu zodpovednosť.

NCC-SK SLOVAKIA CYBERSECURITY COORDINATION CENTRE

ECCC EUROPEAN CYBERSECURITY COMPETENCE CENTRE

Kompetenčné a certifikačné centrum kybernetickej bezpečnosti

Čo je a čo nie je umelá inteligencia?

Čo je AI?

AI je súbor technológií a metód, ktoré umožňujú počítačom a strojom učiť sa z údajov, rozpoznávať vzory, robiť predikcie a vykonávať úlohy, ktoré tradične vyžadovali ľudskú inteligenciu (napr. spracovanie prirodzeného jazyka, rozpoznávanie obrazu, rozhodovacie systémy).

V praxi to znamená, že počítač vám môže pomôcť napríklad:

- Navrhnúť text pozvánky,
- Vybrať vhodného kandidáta na pracovnú pozíciu,
- Skontrolovať gramatiku,
- Zjednodušiť technický výkres,
- Alebo dokonca napísať kód aplikácie.

Čo AI nie je

- Nie je to úplné nahradenie človeka — AI väčšinou dopĺňa ľudskú prácu, nie ju stopercentne nahrádza.
- Nie je neomylná — modely robia chyby (tzv. halucinácie), môžu byť zaujaté a ich výstupy nemusia byť vždy správne.
- Nevie čítať medzi riadkami. Nechápe kontext ako človek, a preto sa môže mýliť.

Prečo musíme používať umelú inteligenciu?

AI mení spôsob práce vo firmách a prináša konkrétne efekty produktivity a efektivity. To znamená, že AI nám výrazne zjednodušuje prácu. Praktické aplikácie zahŕňajú:

- Automatizácia rutinných administratívnych úloh (napr. prepisovanie údajov, sumarizovanie emailov)
- Generovanie textov (správy, e-maily, prezentácie),
- Pomoc pri tvorbe vzorcov v Exceli a spracovaní tabuliek,
- Kontrola gramatiky a štylistiky,
- Predvýber kandidátov (triedenie životopisov),
- Plánovanie trás a optimalizácia logistiky,
- Návrh a zjednodušenie technických výkresov,
- Predikcia správania zákazníkov a personalizácia služieb,
- Generovanie videí a zdrojového kódu.

Tieto možnosti zvyšujú konkurencieschopnosť a umožňujú využiť ľudský čas na úlohy s vyššou pridanou hodnotou. Čím viac úloh AI zvládne, tým viac sa môžeme sústrediť na zmysluplnejšiu prácu. A to je jej hlavná sila.

Spolufinancovaný Európskou úniou

Financované Európskou úniou. Vyjadrené názory a postoje sú názormi a vyhláseniami autora(-ov) a nemusia nevyhnutne odrážať názory a stanoviská Európskej únie. Európska únia za ne nepreberá žiadnu zodpovednosť.

NCC-SK SLOVAKIA CYBERSECURITY COORDINATION CENTRE

ECCC EUROPEAN CYBERSECURITY COMPETENCE CENTRE

Kompetenčné a certifikačné centrum kybernetickej bezpečnosti

Aké sú hlavné hrozby pri používaní AI?

AI je ako nôž – môže rezať chlieb, ale aj ublížiť, ak sa nepoužíva opatrne. Pri adopcii AI sa objavujú nové typy rizík, ktoré si vyžadujú riadené zavádzanie a bezpečnostné kontroly:

- Používanie falošných AI modelov — zamestnanci môžu používať nezabezpečené a neoverené služby, ktoré kradnú alebo logujú zadávané dáta.
- Únik informácií vkladaných do modelu — dáta vložené do verejných modelov (promptov) môžu byť neúmyselne zdieľané alebo uložené poskytovateľom.
- Únik samotného modelu — interné modely s obchodnou hodnotou môžu byť exfiltrované.
- Vytváranie malwaru pomocou AI — modely môžu pomôcť generovať škodlivý kód alebo uľahčiť exploity.
- Tvorba phishingu — AI zrýchľuje tvorbu presvedčivých podvodných správ.
- Útoky na identitu a deepfake — falošné audio/video materiály môžu byť použité na podvody.
- Halucinácie — modely môžu vymýšľať fakty alebo dáta, čo vedie k chybným rozhodnutiam.
- Informačné bubliny — personalizované výstupy môžu utvrdzovať neúplné alebo zaujaté pohľady.
- Úmyselné vyvolanie škodlivých výsledkov — zneužitie promptov na dosiahnutie škodlivých konaní.

Platí nasledovné pravidlo: Nikdy AI nezadávajte informácie, ktoré by ste neposlali cudzincovi.

Praktické príklady nevhodných promptov (čo zamestnanci nesmú robiť)

Do AI nástrojov by sa nikdy nemali vkladať údaje, ktoré by pri úniku spôsobili škodu. Príklady, ktoré predstavujú vysoké riziko:

„Napíš mi pozvánku pre týchto ľudí: Anna Nováková, Peter Kováč…"

- Riziko: Zdieľanie osobných údajov bez ich súhlasu.

„Zhrň mi tieto e-maily od klientov do jedného listu."

- Riziko: AI dostane prístup k súkromnej komunikácii.

„Ako upraviť Excel tak, aby sme zamestnancom od marca znížili plat o 20 %?"

- Riziko: Prezradenie interných informácií o firme.

„Ktorou cestou mám poslať 10 miliónov eur do troch pobočiek?"

- Riziko: AI môže vyzradiť citlivé finančné operácie.

„Kde na mojej adrese Uhrínova 12 sa dajú kúpiť nové zámky?"

- Riziko: Nevedomky priznávate, že vaše dvere nie sú zabezpečené.

„Čo mi hrozí na základe mojej lekárskej správy?"

- Riziko: V dopyte na AI boli prezradené vysoko citlivé medicínske údaje.

Otázky môžu vyzerať nevinne, ale pre kyberzločincov sú cenným zdrojom informácií.

Spolufinancovaný Európskou úniou

Financované Európskou úniou. Vyjadrené názory a postoje sú názormi a vyhláseniami autora(-ov) a nemusia nevyhnutne odrážať názory a stanoviská Európskej únie. Európska únia za ne nepreberá žiadnu zodpovednosť.

NCC-SK SLOVAKIA CYBERSECURITY COORDINATION CENTRE

ECCC EUROPEAN CYBERSECURITY COMPETENCE CENTRE

Kompetenčné a certifikačné centrum kybernetickej bezpečnosti

Ktoré veci by zamestnanci nikdy nemali zadať do AI

- Firemné dôverné údaje — obchodné plány, zastrešujúce stratégie, projekty, cenové ponuky, finančné výkazy pred zverejnením.
- Osobné údaje — rodné čísla, zdravotné záznamy, mená, adresy, emailové adresy, finančné informácie zamestnancov alebo zákazníkov.
- Procesné údaje — interné postupy, bezpečnostné protokoly, nastavenia infraštruktúry.
- Údaje umožňujúce identifikáciu osoby — čo sa týka vás, vašej rodiny alebo domova
- Pre jednotlivcov je nevyhnutné si pripomínať, aby nikdy nezadávali súkromné alebo jedinečné údaje do otvorených (nešifrovaných) modelov.
- Nezabúdajte: AI nie je dôverný poradca. Je to verejný nástroj, ktorý spracúva dáta. A ak sa dáta dostanú do nesprávnych rúk, môžu byť zneužité.

Odporúčané bezpečnostné opatrenia a pravidlá (best practices)

Aby organizácia získala prínosy AI a minimalizovala riziká, odporúčame organizácií implementovať tieto kroky:

- Vytvorte politiku používania AI a školenia, naučte seba aj kolegov, čo je bezpečné zadávať AI a čo nie. Vytvorte si firemné pravidlá používania AI.
- Zaviesť jasnú internú politiku, ktoré AI nástroje sú povolené a ktoré nie. Používajte len schválených poskytovateľov/modely s obchodnými zmluvami (SLA, záruky ochrany dát).
- Nastaviť pravidelné školenia zamestnancov o rizikách (o zadávaní promptov, bezpečnosti vložených citlivých údajov).
- Nasadiť klasifikáciu dát a podporiť ju cez DLP (Data Loss Prevention), obmedzte zadávanie celých dokumentov alebo databáz.
- Každý údaj v organizácií klasifikovať (verejné, interné, dôverné, citlivé).
- Nastaviť DLP pravidlá, ktoré blokujú posielanie citlivých dát do externých AI služieb.
- Preferovať on-premise alebo privátne nasadenia pri kritických dátach.
- Dbať na dobre nastavené prístupové práva a audit.
- Minimalizovať prístup k citlivým modelom a dátam.
- Logovať všetky interakcie s AI systémami a pravidelne auditovať použitie.
- Používať anonymizáciu a pseudonymizáciu údajov.
- Použiť syntetické alebo agregované ukážky namiesto reálnych údajov, ak je to možné.
- Overovať odporúčania a výstupy AI ľudsky, najmä pri rozhodnutiach s finančnými, právnymi alebo zdravotnými dôsledkami.
- Monitorovať modely na príznaky halucinácií alebo skrytého zaujatia.
- Mať pripravený plán reakcie na únik modelu alebo dát z promptov (notifikácia dotknutých, forenzná analýza, právne kroky).
- Pripravte si EXIT plán na odchod z AI riešenia, aby v prípade odchodu bolo zabezpečené vymazanie dát organizácie z AI riešenia.

Spolufinancovaný Európskou úniou

Financované Európskou úniou. Vyjadrené názory a postoje sú názormi a vyhláseniami autora(-ov) a nemusia nevyhnutne odrážať názory a stanoviská Európskej únie. Európska únia za ne nepreberá žiadnu zodpovednosť.

NCC-SK SLOVAKIA CYBERSECURITY COORDINATION CENTRE

ECCC EUROPEAN CYBERSECURITY COMPETENCE CENTRE

Kompetenčné a certifikačné centrum kybernetickej bezpečnosti

Záver

AI nie je „postrachom" kybernetickej bezpečnosti — je to nástroj s obrovským potenciálom, ale zároveň s novými typmi rizík. Správnym riadením, jasnými pravidlami, technickými kontrolami a kultúrou zodpovedného používania môžu organizácie maximalizovať výhody AI a minimalizovať hrozby. Kľúčom je predvídavosť: identifikovať citlivé dáta, kontrolovať ich tok, vzdelávať ľudí a používať len bezpečne spravované AI riešenia.

Spolufinancovaný Európskou úniou

Financované Európskou úniou. Vyjadrené názory a postoje sú názormi a vyhláseniami autora(-ov) a nemusia nevyhnutne odrážať názory a stanoviská Európskej únie. Európska únia za ne nepreberá žiadnu zodpovednosť.

NCC-SK SLOVAKIA CYBERSECURITY COORDINATION CENTRE

ECCC EUROPEAN CYBERSECURITY COMPETENCE CENTRE

Kompetenčné a certifikačné centrum kybernetickej bezpečnosti