Kompetenčné a certifikačné centrum kybernetickej bezpečnosti

NCC-SK
SLOVAKIA CYBERSECURITY COORDINATION CENTRE

KYBERNETICKÁ BEZPEČNOSŤ SA ZAČÍNA ČLOVEKOM

Dôvera, zodpovednosť a bezpečnostná kultúra

Mgr. Ladislav Šnapko
Konferencia TFTU

ČLOVEK NA ZAČIATKU AJ NA KONCI

Pred použitím

Určujeme účel, vyberáme nástroj a rozhodujeme, ktoré údaje mu sprístupníme.

Citlivé podrobnosti zvažujeme ešte pred prvou otázkou.

Počas práce

Overujeme podstatné tvrdenia, zvažujeme kontext a rozhodujeme o použití návrhu.

Kontrolujeme aj podmienky, ktoré návrh vynechal.

Pri výsledku

Chránime súkromie, dôveru a rozhodovanie človeka, ktorého sa odpoveď týka.

Príjemca sa môže podľa našej odpovede rozhodnúť.

AI POMÁHA A ČLOVEK POSUDZUJE

Príprava práce

AI môže navrhnúť osnovu, upraviť jazyk alebo zhrnúť podklad.

Pomoc posudzujeme podľa konkrétnej pracovnej úlohy.

Kontrola významu

Človek preverí fakty, zachovanie podmienok a vhodnosť pre adresáta.

Aj jazyková úprava môže zmeniť význam odpovede.

Kontrola podľa dopadu

Čím vážnejšie následky môže mať omyl, tým dôkladnejšie výsledok preverujeme.

Pri osobnej žiadosti zvažujeme následky odporúčania.

PRESVEDČIVÁ ODPOVEĎ POTREBUJE OVERENIE

Plynulosť nestačí

AI môže presvedčivo uviesť nesprávny údaj alebo vytvoriť neexistujúci zdroj.

Vráťme sa k podkladu

Dôležité tvrdenie overíme v pôvodnom dokumente alebo nezávislom dôveryhodnom zdroji.



Ilustrácia modelovej situácie

Pri súhrne žiadosti preveríme aj vynechané výnimky.

SÚKROMIE CHRÁNIME UŽ PRI ZADANÍ

Iba potrebné údaje

Nevkladáme celý dokument, ak úloha potrebuje len všeobecný alebo fiktívny príklad.

Najprv si ujasníme účel a rozsah potrebných informácií.

Meno nie je jediný identifikátor

Človeka môžu prezradiť aj okolnosti jeho osobného príbehu.

Rozhoduje aj kombinácia pozície, rodiny a udalosti.

Overené prostredie

Používame schválený nástroj a poznáme hranice jeho použitia pre konkrétne údaje.

Oprávnenie čítať dokument samo nestačí na jeho odovzdanie AI.

DÔVERA POTREBUJE PRIMERANÉ OVERENIE

Známy prejav nemusí potvrdiť pravosť

Text, hlas aj obraz môžu pôsobiť dôveryhodne a pritom zavádzať.

AI môže napodobniť aj známu osobu a obvyklý štýl správy.

Citlivú požiadavku potvrdíme

Zmenu účtu alebo prístupu overíme nezávislou cestou cez známy kontakt.

Kontakt uvedený iba v podozrivej správe na overenie nestačí.

Naliehavosť nemení pravidlá

Dohodnutý postup platí aj pri požiadavke vedenia a pri časovom tlaku.

Jednotný postup chráni pracovníka aj samotné vedenie.

ČLOVEK POTREBUJE BEZPEČNÉ PODMIENKY

Jasné pravidlá

Človek vie, ktorý nástroj použiť a aké údaje doň smie vložiť.

Rozlišuje verejné oznámenie od citlivého osobného príbehu.

Čas a podklady

Má možnosť overiť výsledok a pri neistote prácu pozastaviť.

Rýchlosť práce má ponechať priestor na skutočnú kontrolu.

Dostupná pomoc

Pozná konkrétny kontakt a môže sa ozvať bez obavy z ponižovania.

Vecná reakcia na chybu umožní skôr obmedziť jej následky.

ĽUDSKÝ DOHĽAD MUSÍ BYŤ SKUTOČNÝ

Porozumieť a zasiahnuť

Kontrolujúci človek potrebuje podklady, čas a právomoc návrh upraviť alebo odmietnuť.

Jasné miesto zodpovednosti

Vopred určíme, kto preverí obsah a kto rozhodne o odoslaní alebo vykonaní kroku.



Ilustrácia modelovej situácie

Človek môže priznať neistotu a vyžiadať si odbornú pomoc.

ETIKA AI PODĽA COECKELBERGHA

Zodpovednosť voči dotknutému človeku

Mark Coeckelbergh: komu máme vysvetliť dôvody rozhodnutia, ktoré AI ovplyvnila?

Pýtame sa aj na človeka, ktorý musí znášať následky.

Vysvetlenie musí byť zrozumiteľné

Človek potrebuje rozumieť tomu, prečo sme návrh AI prijali a ako sa ho výsledok dotýka.

Pri odpovedi študentovi uvedieme podklady aj dôvod záveru.

Dôvera potrebuje ochotu odpovedať

V našej praxi to znamená vypočuť otázku, priznať neistotu a umožniť preverenie.

Za výsledkom má stáť konkrétna zodpovedná osoba.

OD ETIKY AI K BEZPEČNOSTNEJ KULTÚRE

Hodnoty v každodennej praxi

Inšpirácia Markom Coeckelberghom: hodnoty premietame do návrhu AI aj do jej používania.

Sledujeme, ako nástroj mení zaobchádzanie s ľuďmi.

Zodpovednosť má konkrétne úlohy

V našej organizácii určíme, kto preverí dopad na človeka a kto zabezpečí nápravu.

Preveríme aj potreby človeka, ktorému všeobecná odpoveď nestačí.

Dôvera rastie zo skúsenosti

Ľudia môžu spochybniť nespravodlivý výsledok a bezpečne upozorniť na problém.

Podnet má viesť k odpovedi a podľa zistení aj k zmene postupu.

BEZPEČNOSTNÁ KULTÚRA ŽIJE VO VZŤAHOCH

Otvorene o používaní AI

Ľudia môžu povedať, s čím im nástroj pomáha a kde si nie sú istí.

Rozhovor pomáha odhaliť potreby aj používanie AI mimo pravidiel.

Vedenie ide príkladom

Dodržiava pravidlá a prijíma kontrolu aj pri vlastných výstupoch.

Overovanie je súčasťou kvality práce aj pri návrhu vedúceho.

Z chyby vzniká poučenie

Skúmame rozhodnutie aj podmienky práce a upravujeme príčinu problému.

Ľudia sa dozvedia, čo sa na základe ich podnetu zlepšilo.

PRI CHYBE NAJPRV POMÔŽME RIEŠIŤ SITUÁCIU

Zastavme ďalšie riziko

Neoverenú odpoveď ďalej nešírime a citlivú akciu pozastavíme.

Rozlišujeme návrh textu od akcie, ktorú systém už vykonal.

Zistime presný priebeh

Aké údaje, do ktorej služby a kedy? Čo nástroj skutočne vykonal?

Presné informácie umožnia primerane posúdiť dosah udalosti.

Zapojme pomoc a napravme následky

Podpora preverí dosah. Opravíme chybnú informáciu a upravíme príčinu problému.

Nápravu potrebujú aj príjemcovia už odoslanej chybnej odpovede.

TECHNICKÁ OCHRANA PODPORUJE ČLOVEKA

Schválený nástroj

Organizácia preverí prostredie a podmienky spracovania údajov.

Používateľ má poznať povolený účel a hranice použitia.

Iba potrebné oprávnenia

Prístup k pošte a súborom obmedzíme podľa úlohy a pravidelne ho preverujeme.

Nástroj na návrh textu nemusí mať možnosť meniť záznamy.

Kontrola významných akcií

Návrh odpovede a jej odoslanie sú odlišné kroky. Citlivý úkon potrebuje primerané overenie.

Dostupné záznamy a odborná podpora pomôžu preveriť chybu.

VEDENIE VYTVÁRA PODMIENKY ZODPOVEDNOSTI

Určuje účel a hranice

Pomenuje povolené použitia AI a situácie, ktoré potrebujú ďalšie schválenie.

Rozhoduje aj o tom, kedy treba odborné posúdenie.

Poskytuje zdroje a pomoc

Zabezpečí vhodný nástroj, vzdelávanie, čas na kontrolu a zodpovedné osoby.

Vzdelávanie vychádza z úloh, ktoré pracovníci skutočne riešia.

Reaguje na skúsenosti

Pri zmene nástroja, údajov alebo účelu znovu posúdi riziká a upraví postup.

Pripojenie ďalšieho zdroja môže zmeniť pôvodný dosah AI.

ISO/IEC 42001 PREPÁJA ĽUDÍ A RIADENIE AI

Zodpovednosti v celom procese

Systém manažérstva AI prepája účel, pravidlá, zdroje a priebežné zlepšovanie.

Organizácia sleduje, či dohodnuté opatrenia fungujú v praxi.

Dopad na konkrétneho človeka

Posudzujeme, koho použitie AI zasiahne a čo môže znamenať nesprávna odpoveď alebo sprístupnenie údajov.

Pri citlivej odpovedi zabezpečíme cestu k odbornej pomoci.

Overenie zostáva potrebné

Dobre nastavené riadenie podporuje dôveru. Konkrétny výstup stále potrebuje primeranú kontrolu.

Certifikácia sama nepotvrdzuje správnosť každej odpovede AI.

EU AI ACT: OCHRANA ČLOVEKA PODĽA RIZIKA

O rozsahu ochrany rozhoduje účel AI a jej dopad na človeka.

Neprijateľné riziko	Niektoré manipulatívne použitia a sociálne skórovanie spĺňajúce zákonné podmienky sú zakázané.
Vysoké riziko	Vybrané použitia pri nábore či prijímaní do školy potrebujú riadenie rizík, dohľad a kybernetickú ochranu.
Transparentnosť	Pri vybraných použitiach treba oznámiť interakciu s AI alebo umelý pôvod obsahu, napríklad deepfake.
Minimálne riziko	Spamové filtre či niektoré hry. Naďalej zohľadňujeme ochranu údajov a ďalšie uplatniteľné pravidlá.

AI gramotnosť a jasné zodpovednosti podporujú bezpečnostnú kultúru.

Rozsah povinností závisí aj od postavenia organizácie a termínu ich uplatňovania.

ZAČNIME JEDNOU KONKRÉTNOU ÚLOHOU

Poznajme skutočnú prax

Vyberme jednu úlohu a zistime, aký nástroj a údaje pri nej ľudia používajú.

Rozhovor s pracovníkom ukáže užitočnosť aj jeho neistotu.

Dohodnime bezpečný postup

Určime hranice vstupu, zodpovednú osobu a spôsob overenia výsledku.

Pracovník vie, čo môže urobiť sám a kedy požiada o pomoc.

Vyskúšajme a upravme

Na fiktívnom príklade preverme aj chybu a dostupnosť pomoci. Odstráňme zistené prekážky.

Sledujeme, či človek nájde podklad a kontakt na podporu.

ZMYSLOM BEZPEČNOSTI JE OCHRANA ČLOVEKA

Dôvera má konkrétny základ

Rozumieme hraniciam nástroja a overujeme to, o čo sa má niekto oprieť.

Neistotu vieme pomenovať a dôležitý záver vysvetliť.

Zodpovednosť potrebuje podmienky

Človek má podklady, právomoc rozhodnúť a dostupnú pomoc.

Organizácia vytvára priestor na kontrolu aj nápravu chyby.

Na konci je ďalší človek

Bezpečnosť sa začína človekom a jej zmysel sa napĺňa v ochrane človeka.

Jeho súkromie a dôvera dávajú bezpečnosti konkrétny zmysel.

Kompetenčné a certifikačné centrum kybernetickej bezpečnosti

Ďakujem za pozornosť

Mgr. Ladislav Šnapko

Certifikovaný audítor kybernetickej bezpečnosti

Certifikovaný manažér kybernetickej bezpečnosti