ROLA A VÝZNAM UMELEJ INTELIGENCIE V PRAXI SOCIÁLNEJ PRÁCE NOVÁ PARADIGMA V KLINICKOM ROZHODOVANÍ

Petronela ŠEBESTOVÁ – Jana LAŠČIAKOVÁ

THE ROLE AND SIGNIFICANCE OF ARTIFICIAL INTELLIGENCE IN SOCIAL WORK PRACTICE A NEW PARADIGM IN CLINICAL DECISION-MAKING

Abstrakt: Článok sa zaoberá komplexnou a ambivalentnou rolou umelej inteligencie -Artificial Intelligence (AI) v oblasti sociálnej práce. Prostredníctvom analyticko-syntetickej metódy skúma potenciálne prínosy AI, ako sú optimalizácia administratívy, podpora klinického rozhodovania a prediktívna analýza. Komparatívnou analýzou konfrontuje tieto príležitosti s inherentnými rizikami, menovite algoritmickou zaujatosťou, eróziou terapeutického vzťahu, problémom transparentnosti a etickou zodpovednosťou. Záverečná syntéza argumentuje, že dichotómia „nástroj verzus hrozba“ je falošná. Namiesto toho navrhuje koncepčný rámec pre implementáciu AI ako augmentačnej technológie v modeli „Human-in-the-Loop“, kde konečné rozhodnutie a etická zodpovednosť zostávajú pevne v rukách sociálneho pracovníka. Human-in-the-Loop“ (skratka HITL), alebo „človek v slučke“, je model spolupráce, kde umelá inteligencia (AI) a človek pracujú v tíme na vyriešení problému. Nie to ako súboj „človek verzus stroj“, ale ako partnerstvo, kde každý robí to, v čom je najlepší. Umelá inteligencia (AI) je skvelá v rýchlom spracovaní obrovského množstva dát a hľadaní vzorcov. Človek je nenahraditeľný v chápaní kontextu, irónie, etiky a v riešení neštandardných, nejednoznačných situácií. Human-in-the-Loop spája tieto dva svety do jedného cyklu alebo „slučky“.

Kľúčové slová: umelá inteligencia, sociálna práca, etika, algoritmická zaujatosť, prediktívna analýza, augmentovaná inteligencia.

Abstract: This article addresses the complex and ambivalent role of artificial intelligence (AI) in the field of social work. Using an analytical-synthetic method, the paper examines the potential benefits of AI, such as the optimization of administrative tasks, clinical decision support, and predictive analytics. Through comparative analysis, these opportunities are juxtaposed with inherent risks, namely algorithmic bias, the erosion of the therapeutic

relationship, challenges of transparency, and ethical accountability. The final synthesis argues that the "tool versus threat" dichotomy is a false one. Instead, the paper proposes a conceptual framework for implementing AI as an augmentation technology within a 'Human-in-the-Loop' (HITL) model. This model is framed as a collaborative partnership that leverages the respective strengths of both AI (rapid data processing and pattern recognition) and human professionals (contextual understanding, ethical reasoning, and managing ambiguity). Within this framework, final decision-making and ethical responsibility remain firmly vested in the human social worker.

Keywords: artificial intelligence, social work, ethics, algorithmic bias, predictive analytics, augmented intelligence.

Úvod

Nástup umelej inteligencie predstavuje technologický posun porovnateľný s priemyselnou revolúciou. Zatiaľ čo jej implementácia v sektoroch ako napr. financie či logistika je relatívne priamočiara, avšak jej vstup do oblastí definovaných ľudskou interakciou, empatiou a etickým úsudkom, akou je sociálna práca, vyvoláva fundamentálne otázky. Sociálna práca, ako profesia zameraná na podporu zraniteľných jedincov a presadzovanie sociálnej spravodlivosti, stojí na križovatke. Na jednej strane AI sľubuje bezprecedentnú efektivitu a dátami podložené intervencie, na druhej strane hrozí zavedením systémovej diskriminácie a dehumanizáciou pomáhajúceho vzťahu. Tento článok si kladie za cieľ dekonštruovať túto dichotómiu a ponúknuť kritickú analýzu potenciálu a rizík AI, ktorá vyústi do syntézy eticky udržateľného modelu jej využitia.

1 Príležitosti a potenciál umelej inteligencie v sociálnej práci

Potenciál umelej inteligencie v sociálnej práci možno segmentovať do štyroch kľúčových oblastí, ktorými sú:

1.1 Prvý segment - Optimalizácia administratívnych a procesných úloh - sociálni pracovníci sú chronicky zaťažení administratívou, ktorá im uberá čas na priamu prácu s klientom. Nástroje AI využívajúce spracovanie prirodzeného jazyka (NLP) dokážu automatizovať transkripciu rozhovorov, sumarizovať kazuistiky, analyzovať dokumentáciu a identifikovať kľúčové informácie. Systémy riadenia zdrojov môžu na základe profilu klienta

automaticky navrhovať najvhodnejšie dostupné služby (napr. typ ubytovania, forma terapie), čím dramaticky zvyšujú efektivitu.

Pre pregnantnejšie pochopenie uvedieme kazuistiku na prvý kľúčový segment. Klient príde do Komunitného centra, v ktorom sa podľa § 24d zákona 448/2008 o sociálnych službách poskytuje fyzickej osobe v nepriaznivej sociálnej situácii aj sociálne poradenstvo. Klient sa ocitol v zložitej situácii. Nedávno stratil bývanie, má cukrovku, ktorú si nepravidelne lieči, a potrebuje si vybaviť nové doklady, aby si mohol hľadať prácu. Sociálny pracovník podľa tradičného postupu, teda bez použitia AI nadviaže s klientom rozhovor a súčasne sa snaží písať poznámky, aby zachytil všetky dôležité informácie. Musí neustále prepínať medzi aktívnym počúvaním (budovaním dôvery) a písaním. Po prepísaní poznámok z celého rozhovoru do počítača musí vytvoriť kazuistiku, aby sa v prípade vedel rýchlo zorientovať a aby boli všetky potreby klienta zaznamenané. Musí si spomínať aj na detaily, ktoré si nestihol zapísať. Po vytvorení kazuistiky sa opäť vracia k textu a manuálne vyberá najdôležitejšie informácie a súčasne vytvára identifikáciu relevantných potrieb pre klienta (nocľaháreň, diabetické ambulancie pre pacientov bez bydliska, kontakty na pracovné agentúry, úradné hodiny oddelení na vybavenie dokladov, strediská hygieny a pod.). Tento postup hľadania zdrojov a informácií zaberá sociálnemu pracovníkovi veľa času a okrem toho sa dostáva do stresovej situácie, lebo čas venovaný administratíve mohol využiť na sociálnu pomoc klientovi, ktorý sa taktiež ocitol v nepriaznivej sociálnej situácii. (Goldkind, Lauri a John G. MCnutt (eds.). 2021).

Pri pracovnom postupe s **umelou inteligenciou (AI)** sa môžu pracovné výkony optimalizovať. Sociálny pracovník má nahrávacie zariadenie a so súhlasom klienta je celý rozhovor nahrávaný a celá pozornosť sociálneho pracovníka je venovaná iba klientovi. Sociálny pracovník sa dostáva do pozície kouča, a je plne prítomný, udržiava očný kontakt, je empatický a **buduje vzťah**. Klient je spokojný, lebo cíti pochopenie, pomoc a východisko na zmiernenie resp. odstránenie svojich problémov. Optimalizácia pracovných výkonov prostredníctvom AI spočíva v automatickej transkripcii, pri ktorej je zvuková nahrávka spracovaná prostredníctvom neurolingvistického programovania **(NLP)** prevedie celý rozhovor na písaný text. Potom nastáva **inteligentná sumarizácia**. AI softvér následne analyzuje celý prepis a automaticky vygeneruje štruktúrovaný súhrn v podobe: kto je klient, aká je jeho situácia, aké sú jeho hlavné problémy a aké si stanovil ciele. AI zároveň v texte podčiarkne a otaguje (priradí) kľúčové potreby. Systém na základe týchto tagov automaticky prehľadá aktuálnu databázu dostupných služieb a zdrojov a na obrazovke sa zobrazí panel s návrhmi na uvedené problémy a potreby klienta. AI v tomto príklade neurobila žiadne rozhodnutie za

sociálneho pracovníka, len mu urobila "asistenta" na tú najnudnejšiu a časovo najnáročnejšiu prácu, čím mu uvoľnila ruky, aby mohol robiť to, čo je naozaj dôležité – pracovať s človekom.

1.2 Druhý segment - Prediktívna analýza a včasná intervencia - je jedna z najsilnejších, ale aj najkontroverznejších aplikácií AI v sociálnej práci. Algoritmy strojového učenia môžu analyzovať rozsiahle datasety (napr. demografické údaje, zdravotné záznamy, históriu kontaktov so sociálnymi službami) s cieľom identifikovať jedincov alebo rodiny s vysokým rizikom budúcich problémov, ako je napríklad zanedbávanie detí, bezdomovectvo či recidíva. Cieľom nie je deterministicky predpovedať budúcnosť, ale alokovať obmedzené zdroje na preventívne opatrenia tam, kde je to najviac potrebné. Na konkrétnom príklade, ktorým je prevencia krízovej situácie u osamelo žijúceho seniora budeme analyzovať druhý segment AI v sociálnej práci. Východiskovou situáciou budú údaje o poskytovaní sociálnych služieb seniorom v obci. Dôležité je, že ide o dáta, ktorými už obec disponuje. Sociálny pracovník si vytvorí model, ktorý analyzuje historické dáta a hľadá varovné signály a vzorce, ktoré predchádzali krízovým situáciám v minulosti. Model môže obsahovať údaje napr. o systéme **donášky obedov** (evidencia pravidelnosti odberu, evidencia náhlych odhlásení, alebo opakovaného neprevzatia obeda a pod). Ďalším modelom môže byť d**ochádzkový systém do denných centier pre seniorov (z**áznamy o tom, ako často senior centrum navštevuje, či nedošlo k náhlemu a dlhodobému prerušeniu návštev pod.). V rámci **opatrovateľskej služby** môže byť vytvorený model, ktorý bude evidovať dáta o tom, či senior v poslednom čase odmietol plánovanú návštevu opatrovateľky, alebo je v danom časovom rozpätí poskytnutia služby neprítomný. Po analýze dát stoviek prípadov za posledné roky AI identifikuje resp. vytvorí **vysoko rizikový vzorec správania**. Napríklad ak senior (vek 80+), ktorý pravidelne odoberal obedy, náhle zruší ich odber na viac ako týždeň a zároveň v tom istom mesiaci prestane chodiť do denného centra, ktoré predtým navštevoval aspoň 2x týždenne, existuje 90% pravdepodobnosť, že do 30 dní u neho nastane krízová situácia z rôznych dôvodov: pád, dehydratácia, vážne zhoršenie zdravotného stavu. Systém AI na základe zhody s rizikovým vzorcom automaticky vygeneruje **dôverný alert** (signál, ktorý upozorňuje na dôležitú udalosť, hrozbu alebo krízu), v našom prípade pre terénnu sociálnu pracovníčku obce a odporučí prioritný kontakt aj s určením času.

1.3 Tretí segment - Podpora klinického rozhodovania - klinické rozhodovanie v sociálnej práci je systematický myšlienkový proces, ktorý sociálny pracovník používa na zhodnotenie situácie klienta, stanovenie cieľov a výber najvhodnejších a najefektívnejších postupov (intervencií) na pomoc tomuto klientovi. Nie je to len o "pocite" alebo "intuícii". Je to

profesionálna zručnosť, ktorá spája teóriu, dôkazy, etiku a praktické skúsenosti. V medicíne sa klinické rozhodovanie zameriava hlavne na chorobu a jej liečbu. Cieľom je vyliečenie alebo manažment ochorenia. V sociálnej práci a psychológii je záber širší. Zameriava sa na človeka v jeho životnej situácii (bio-psycho-eduko-spirit-soc. model). Problémy sú často komplexnejšie (chudoba, trauma, závislosť, potreby, vzdelávanie, zdravie) a cieľom nie je vždy "vyliečenie", ale skôr zlepšenie fungovania, stabilizácia alebo posilnenie klienta. V kontexte sociálnej práce sa "klinické" nevzťahuje na nemocnicu alebo kliniku. Odkazuje na prácu, ktorá je priama - týka sa priameho kontaktu a práce s klientmi (jednotlivcami, rodinami, skupinami). Systematická- postupuje sa podľa logických krokov, nie chaoticky. Odborná - je založená na teóriách sociálnej práce, psychológie, sociológie a iných overených postupoch. Cielená - smeruje k dosiahnutiu konkrétnej pozitívnej zmeny v živote klienta. Klinické rozhodovanie je teda cyklický proces, ktorý má niekoľko fáz.

Prvá fáza - zhodnotenie:

***Čo sa deje*?** Sociálny pracovník aktívne zbiera informácie. Nejde len o vypočutie si problému. Využíva rozhovory, pozorovanie, štúdium dokumentácie, dotazníky a niekedy aj štandardizované hodnotiace nástroje.

Príklad: Pracovník v zariadení pre seniorov nielen počúva, že klient sa "cíti osamelo", ale zisťuje, ako často má návštevy, aké sú jeho záujmy, aký je jeho zdravotný stav a či nemá príznaky depresie.

Druhá fáza - analýza a syntéza:

Prečo sa to deje? Sociálny pracovník spája získané informácie do zmysluplného celku. Identifikuje silné stránky klienta, rizikové faktory, vzorce správania a príčiny problémov. V tejto fáze využíva teoretické znalosti (napr. teóriu pripútania , systémovú teóriu).

Príklad: Pracovník zistí, že osamelosť klienta nesúvisí len s nedostatkom návštev, ale aj s jeho zhoršenou mobilitou (rizikový faktor je začínajúca artritída).

Čo s tým urobíme? Na základe analýzy terapeut spolu s klientom formuluje ciele a zvažuje rôzne možnosti riešenia. Ktorá intervencia bude najefektívnejšia a najvhodnejšia pre tohto konkrétneho klienta v jeho unikátnej situácii?

Príklad: Namiesto jednoduchého "zabezpečiť viac návštev" terapeut navrhne kombináciu krokov: zabezpečiť fyzioterapiu na zvládanie artritídy, zapojiť klienta do klubovej činnosti priamo v zariadení a dohodnúť pravidelné videohovory s rodinou.

Tretia fáza je implementácia - realizácia:

***Poďme to urobiť*.** Vybrané postupy sa realizujú v praxi. Terapeut poskytuje poradenstvo, sprostredkúva služby, vedie rodinné stretnutia atď.

Štvrtá fáza - vyhodnotenie:

Funguje to? Pracovník priebežne sleduje, či zvolené postupy prinášajú očakávané výsledky. Je situácia klienta lepšia? Ak nie, prečo? Je potrebné plán zmeniť? Tým sa kruh uzatvára a začína sa nové zhodnotenie.

Príklad: Po mesiaci terapeut zisťuje, že klient navštevuje klub len sporadicky. Zisťuje prečo a upravuje plán.

V prípade, že sociálny pracovník - terapeut využije AI pri podpore klinického rozhodovania môžeme konštatovať uvedené rozdiely:

Aspekt	**Tradičný prístup bez použitia AI**	**Prístup s použitím AI**
Objektivita	Vysoká miera subjektivity, závislosť od skúseností.	Vyššia objektivita vďaka štandardizovanému nástroju.
Riziko chyby	Možnosť prehliadnuť dôležité faktory.	Minimalizácia rizika vďaka automatickým varovaniam a kontrolám.
Konzistentnosť	Rozhodnutia rôznych pracovníkov sa môžu výrazne líšiť.	Zabezpečenie konzistentnosti a rovnakého prístupu
Efektivita	Zapisovanie, analýza a plánovanie sú oddelené procesy.	Integrácia procesov, automatizácia vyhodnotenia a návrhov šetrí čas.
Využitie EBP	Závisí od znalostí konkrétneho pracovníka.	Systém aktívne ponúka postupy založené na dôkazoch a legislatíve.
Dokumentácia	Menej štruktúrovaná, ťažšie sledovateľná.	Prehľadná, štruktúrovaná a ľahko auditovateľná dokumentácia.

AI môže slúžiť ako asistenčný nástroj, ktorý sociálnemu pracovníkovi- terapeutovi poskytne syntézu relevantných informácií. Na základe analýzy dát konkrétneho klienta a porovnania s tisíckami anonymizovaných prípadov a výsledkov výskumov, môže systém navrhnúť najúčinnejšie, dôkazmi podložené (evidence-based) intervenčné postupy. Komparatívne, ako rádiológ využíva AI na detekciu anomálií na snímkach, sociálny pracovník - terapeut by mohol využiť AI na identifikáciu skrytých vzorcov v správaní klienta. Tento príklad ukazuje, že systém na podporu klinického rozhodovania nenahrádza odborný úsudok sociálneho pracovníka-terapeuta, ale slúži ako silný nástroj, ktorý ho usmerňuje, znižuje riziko zlyhania ľudského faktora a pomáha robiť informovanejšie a lepšie obhájiteľné rozhodnutia v prospech klienta.

1.4 Vzdelávanie a supervízia: AI môže vytvárať pokročilé simulačné prostredia, kde si študenti sociálnej práce môžu nacvičovať krízovú intervenciu s virtuálnymi klientmi

(chatbotmi), ktorí reagujú na základe rôznych psychologických modelov. V supervízii môže (s informovaným súhlasom klienta) AI analyzovať prepisy sedení a upozorniť supervízora na verbálne aj neverbálne signály, ktoré mohol supervidovaný pracovník prehliadnuť. Zavedenie umelej inteligencie (AI) do sociálnej práce paradoxne nezničuje, ale naopak, extrémne zvyšuje dôležitosť a mení **význam vzdelávania a supervízie.** Už to nie je len o tom, ako pomôcť klientovi. Je to aj o tom, ako správne a bezpečne zaobchádzať s novým, veľmi mocným nástrojom.

Nový význam VZDELÁVANIA v ére AI. Vzdelávanie sa už nemôže sústrediť len na sociálne teórie a metódy práce. Musí z pracovníkov v sociálnej práci urobiť inteligentných a kritických používateľov technológie.

- ⇨ **Od znalosti k gramotnosti:**
  - **Predtým:** Pracovník sa musel naučiť zákony a postupy.
  - **Teraz s AI:** Pracovník sa musí naučiť aj **technologickú gramotnosť**. Nemusí vedieť programovať, ale musí rozumieť základným princípom: Čo je AI? Ako sa "učí"? Aké sú jej obmedzenia? Prečo sa môže mýliť?
- ⇨ **Tréning kritického myslenia:**
  - **Predtým:** Vzdelávanie učilo, ako si vytvoriť vlastný odborný názor.
  - **Teraz s AI:** Vzdelávanie musí aktívne trénovať pracovníkov, aby spochybňovali výstupy AI. Ak AI asistent navrhne pre klienta 5 riešení, pracovník sa musí pýtať: *Prečo práve tieto? Aké dáta AI použila? Nechýba v jej návrhu nejaký dôležitý ľudský faktor?*
- ⇨ **Nová etika:**
  - **Predtým:** Etika sa zaoberala vzťahom medzi pracovníkom a klientom.
  - **Teraz s AI:** Vzdelávanie musí pokrývať úplne nové etické otázky: Ako chrániť citlivé dáta klienta, ktoré spracúva AI? Čo robiť, ak viem, že AI môže byť voči niektorým skupinám ľudí zaujatá (tzv. algoritmická zaujatosť)? Kto je zodpovedný, ak AI navrhne zlý postup?

Vzdelávanie už nie je len o tom, aby mal pracovník v hlave "mapu". Je o tom, aby sa naučil používať "GPS" (AI), ale zároveň neprestal sledovať cestu a okolie vlastnými očami.

Nový význam SUPERVÍZIE v ére AI

Supervízia sa stáva kľúčovým "ľudským kontrolným bodom" a "etickým kompasom" v procese, kde spolupracuje človek a stroj.

- ⇨ **Ochrana pred "automatizačnou slepotou":**
  - Ľudia majú tendenciu slepo dôverovať technológiám. Ak počítač povie, že rodina je "vysoko riziková", pracovník to môže automaticky prijať bez hlbšieho zamyslenia.

- Supervízor je ten, kto kladie kľúčovú otázku: ***"AI ti dala tento výsledok, ale čo hovorí tvoj odborný cit a tvoja skúsenosť? Sedí ti to? Poďme sa na to pozrieť spolu."*** Supervízia je priestor, kde sa overuje zhoda medzi dátami z AI a ľudským úsudkom.

⇨ **Riešenie nových etických dilem:**

- Čo ak AI navrhne najefektívnejšie, ale necitlivé riešenie? Čo ak odporučí napr. odobrať dieťa z rodiny na základe štatistického rizika, aj keď pracovník cíti, že rodina má potenciál na zmenu?
- Tieto nové, zložité dilemy sa musia riešiť v bezpečnom priestore supervízie. Supervízor pomáha pracovníkovi nájsť rovnováhu medzi "chladnou" efektivitou stroja a ľudskosťou a empatiou.

⇨ **Integrácia, nie nahradenie:**

- Supervízia pomáha pracovníkovi ujasniť si, ako používať AI ako nástroj, a nenechať sa ním pohltiť. Pomáha mu formulovať otázky pre AI, interpretovať jej odpovede a správne ich "preložiť" do reálneho plánu pomoci pre klienta.

3 Hrozby a etické dilemy

Prísľub efektivity je vyvážený závažnými rizikami, ktoré priamo ohrozujú základné hodnoty sociálnej práce.

3.1 Algoritmická zaujatosť (Bias) a systémová diskriminácia: AI modely sa učia z historických dát. Ak tieto dáta odrážajú existujúce spoločenské predsudky (napr. nadmernú kontrolu menšinových komunít políciou, alebo sociálnymi službami), AI sa tieto predsudky nielen naučí, ale ich aj zosilní a zafixuje do podoby objektívne sa tváriaceho „rizikového skóre“. **Komparatívne:** Zatiaľ čo ľudský predsudok je individuálny a napadnuteľný, algoritmický predsudok je systematický, škálovateľný a skrytý za závojom technologickej neutrality. Namiesto nástroja spravodlivosti sa tak AI môže stať nástrojom automatizovanej opresie. (Eubanks, Virginia. 2018).

3.2 Erózia terapeutického vzťahu a dehumanizácia: Jadrom sociálnej práce je autentický ľudský vzťah postavený na dôvere, empatii a porozumení neverbálnej komunikácii. Žiadny algoritmus nedokáže replikovať ľudskú schopnosť intuitívneho úsudku a kontextuálneho pochopenia jedinečného životného príbehu. Nadmerné spoliehanie sa na AI v diagnostike a rozhodovaní degraduje klienta na súbor dát a sociálneho pracovníka na operátora systému. Tým sa vytráca samotná podstata pomáhajúceho vzťahu.

3.3 Problém transparentnosti („Black Box“) a zodpovednosti: Mnohé pokročilé modely strojového učenia (napr. hlboké neurónové siete) fungujú ako „čierna skrinka“. Dokážu

vygenerovať výsledok (napr. rizikové skóre), ale nie sú schopné zrozumiteľne vysvetliť, na základe akých premenných a s akou váhou k nemu dospeli. Kto je potom zodpovedný za chybné rozhodnutie? Sociálny pracovník, ktorý sa riadil odporúčaním AI? Agentúra, ktorá systém zakúpila? Alebo programátor, ktorý model vytvoril? Táto difúzia zodpovednosti je v príkrom rozpore s princípom profesionálnej zodpovednosti. Black box v AI je systém, ktorý dáva odpovede bez vysvetlení. To je v citlivých oblastiach ako sociálna práca nebezpečné, a preto je nevyhnutná ľudská kontrola. (Eubanks, Virginia. 2018).

3.4 Ochrana súkromia a dátová bezpečnosť: Sociálna práca operuje s najcitlivejšími osobnými údajmi. Centralizácia týchto dát do systémov AI vytvára masívny cieľ pre kybernetické útoky a riziko ich zneužitia na komerčné či politické účely, čo by malo pre klientov devastujúce následky.

4 Od dichotómie k modelu „Augmented Social Worker - „rozšírený“ alebo „vylepšený“ sociálny pracovník“

Je koncept, ktorý neopisuje robota, ale človeka – profesionála, ktorého prirodzené schopnosti sú posilnené a rozšírené pomocou technológie, predovšetkým umelej inteligencie (AI). Nie je to o nahradení človeka, ale o vytvorení synergie alebo partnerstva medzi človekom a strojom. Polemika „nástroj verzus hrozba“ je zavádzajúca, pretože implikuje, že AI je autonómny aktér. AI je však technológia, ktorej dopad je definovaný spôsobom jej implementácie a riadenia. Riešením nie je ani naivné prijatie, ani technofóbne odmietnutie, ale vytvorenie modelu augmentovanej inteligencie, kde AI slúži výlučne na posilnenie, nie nahradenie ľudských schopností. Navrhovaný model **„Human-in-the-Loop“ (človek v rozhodovacom procese)** je založený na nasledujúcich princípoch:

- **AI ako asistent, nie ako sudca:** AI môže spracovať dáta a ponúknuť pravdepodobnostné scenáre a odporúčania, ale finálny úsudok, interpretácia a rozhodnutie sú neodňateľnou a výlučnou doménou ľudského profesionála.
- **Maximalizácia transparentnosti:** Prednosť musia mať modely, ktorých rozhodovacie procesy sú interpretovateľné (tzv. Explainable AI - XAI). Explainable AI je nevyhnutným krokom k zodpovednému a etickému využívaniu umelej inteligencie v dôležitých oblastiach, ako je medicína, súdnictvo alebo sociálna práca. Premieňa AI z tajomnej a potenciálne nebezpečnej „čiernej skrinky“ na transparentného a

dôveryhodného partnera pre profesionálov, ako je náš **„Augmented Social Worker“** Sociálny pracovník musí rozumieť, prečo systém niečo odporúča.

- ⇨ **Etická matica ako podmienka implementácie:** Pred nasadením akéhokoľvek AI nástroja musí prebehnúť prísny etický audit zameraný na identifikáciu a mitigáciu rizika zaujatosti, diskriminácie a ochrany súkromia. Mitigácia rizika je proces, pri ktorom sa prijímajú konkrétne kroky na zníženie pravdepodobnosti výskytu rizika alebo na zmiernenie jeho negatívnych následkov, ak už nastane. Cieľom nie je vždy riziko úplne odstrániť (čo je často nemožné), ale znížiť ho na prijateľnú úroveň. Mitigácia je len jednou zo štyroch hlavných stratégií manažmentu rizika. (D'Agostino, Maria G. G. T. C. a Eglė Butkevičienė (eds.). 2023)

Zmierenie (Mitigácia): Zníženie pravdepodobnosti alebo dopadu rizika.

Vyhýbanie sa (Avoidance): Úplné odstránenie rizika tým, že sa vyhneme aktivite, ktorá ho spôsobuje.

Akceptácia (Acceptance): Vedomé prijatie rizika a jeho možných následkov, často preto, lebo náklady na mitigáciu by boli príliš vysoké.

Prenesenie (Transference): Prenesenie finančných následkov rizika na tretiu stranu.

Mitigácia rizika v kontexte AI neznamená nepoužívať túto technológiu (to by bola stratégia vyhýbania sa). Znamená to používať ju múdro a zodpovedne – s vedomím jej slabých stránok a so zavedenými „ľudskými poistkami“, aby sa maximalizovali jej prínosy a minimalizovali potenciálne škody.

Prehľad rizík AI v sociálnej práci a stratégie ich mitigácie

1. Riziko: Bias a diskriminácia

- **Stratégie na zmiernenie:**
  - Vzdelávanie pracovníkov, aby vedeli bias rozpoznať.
  - Supervízia ako priestor na kritické zhodnotenie výstupov AI.
  - Pravidelný audit dát, na ktorých sa AI trénuje, a ich čistenie od predsudkov.
  - Využívanie Explainable AI (XAI), ktorá vysvetľuje svoje rozhodnutia. (Rudin, Cynthia. 2019).

2. Riziko: Chyby „čiernej skrinky“ (Black Box)

- **Stratégie na zmiernenie:**

- Implementácia modelu Human-in-the-Loop (HITL), kde finálnu kontrolu a rozhodnutie robí vždy človek.
- Stanovenie jasných pravidiel, kedy je ľudský zásah povinný.
- Nikdy nepoužívať AI na plne autonómne rozhodnutia s vážnymi následkami pre klienta.

3. Riziko: Dehumanizácia práce

- **Stratégie na zmiernenie:**
  - Tréningy a supervízia zamerané na dôležitosť udržania ľudského vzťahu s klientom.
  - Jasne definovať rolu AI len ako asistenta na administratívne a analytické úlohy, nie ako náhradu za rozhovor.
  - Čas ušetrený vďaka AI alokovať priamo na predĺženie času stráveného s klientmi.

4. Riziko: Porušenie súkromia a zneužitie dát

- **Stratégie na zmiernenie:**
  - Dôsledné používanie anonymizovaných a pseudonymizovaných dát.
  - Prísne technické zabezpečenie systémov a dodržiavanie legislatívy (napr. GDPR).
  - Transparentná komunikácia s klientmi o tom, ako a prečo sa ich dáta používajú. (Rudin, Cynthia. 2019.)

5 Záver a odporúčania

Umelá inteligencia nie je ani utopickým riešením, ani dystopickou hrozbou. Je to mocný zosilňovač, ktorý bude amplifikovať zámery a hodnoty, s ktorými ho budeme implementovať. Pre sociálnu prácu to znamená, že ak bude AI nasadená v rámci systémov zameraných len na efektivitu a kontrolu, bude zosilňovať nerovnosť a dehumanizáciu. Ak však bude vyvíjaná a riadená v súlade s etickými princípmi profesie, môže zosilniť schopnosť sociálnych pracovníkov poskytovať kvalitnejšiu a dostupnejšiu pomoc.

Odporúčania pre prax, vzdelávanie a politiku:

- **Vzdelávanie:** Do kurikula štúdia sociálnej práce je nevyhnutné zaradiť predmety zamerané na digitálnu gramotnosť, kritické hodnotenie technológií a etiku AI.
- **Regulácia:** Je potrebné vytvoriť jasné legislatívne a etické rámce pre používanie AI v sociálnych službách, vrátane nezávislého auditu algoritmov.

- **Participatívny dizajn:** Sociálni pracovníci a klienti musia byť aktívne zapojení do procesu navrhovania a testovania AI nástrojov, aby sa zabezpečila ich relevancia a bezpečnosť.

Konečným cieľom nie je vytvoriť umelého sociálneho pracovníka, ale použiť umelú inteligenciu na to, aby sme posilnili a podporili toho skutočného. V tom spočíva najväčšia príležitosť a zároveň najväčší záväzok pre budúcnosť sociálnej práce ako vedy, ale aj profesie.

ZOZNAM BIBLIOGRAFICKÝCH ODKAZOV

BUOLAMWINI, Joy a Timnit GEBRU. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In: *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*. New York: PMLR, vol. 81, s. 77-91.

D'AGOSTINO, Maria G. G. T. C. a Eglė BUTKEVIČIENĖ (eds.). 2023. *The Public Sector AI Handbook*. Cham: Palgrave Macmillan. ISBN: 978-3031191833. DOI: 10.1007/978-3-031-19184-0.

EUBANKS, Virginia. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press. 272 s. ISBN 978-1250074317.

GOLDKIND, Lauri a John G. MCNUTT (eds.). 2021. *Social Work and the Digital Age*. New York: Oxford University Press. 400 s. ISBN: 978-0197514110.

O'NEIL, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown. 320 s. ISBN: 978-0553418811.

RUDIN, Cynthia. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. In: *Nature Machine Intelligence*. Vol. 1, no. 5, s. 206–215. ISSN: 2522-5839. DOI: 10.1038/s42256-019-0048-x.

KONTAKT

Doc. PhDr. Petronela Šebestová, PhD., univer. prof., MPH
Vysoká škola Danubius
petronela.sebestova@gmail.com

PhDr. Jana Laščiaková, PhD.
Vysoká škola Danubius
Jana.Lasciakova@vsdanubius.sk