UMELÁ INTELIGENCIA

SOCIÁLNE A ETICKÉ PROBLÉMY

Zostavovatelia
DILBAR ALIEVA, JURAJ SCHENK,
MARTIN TAKÁČ

UNIVERZITA KOMENSKÉHO
V BRATISLAVE

UMELÁ INTELIGENCIA
SOCIÁLNE A ETICKÉ PROBLÉMY

DILBAR ALIEVA, JURAJ SCHENK,
MARTIN TAKÁČ
zostavovatelia

2024
UNIVERZITA KOMENSKÉHO V BRATISLAVE

Zborník vznikol s podporou Centra pre umelú inteligenciu – AIslovakIA a projektu TERAIS (č. 101079338) v rámci programu Horizont Európa.

Zostavovatelia
doc. PhDr. Dilbar Alieva, CSc.
prof. PhDr. Juraj Schenk, CSc.
doc. RNDr. Martin Takáč, PhD.

Recenzenti
prof. RNDr. Jiří Pospíchal, DrSc.
Mgr. Marek Mathias, PhD.

https://stella.uniba.sk/texty/FMFI_AST_ai-socialne_eticke_problemy.pdf

Ilustrácie na obálke
doc. RNDr. Martin Takáč, PhD. – s použitím nástroja Stable Diffusion (images created through Stable Diffusion Online are fully open source, explicitly falling under the CC0 1.0 Universal Public Domain Dedication)
Ing. Jakub Suchý

Vydavateľ
Univerzita Komenského v Bratislave, 2024

ISBN 978-80-223-5793-7 (tlač)
ISBN 978-80-223-5794-4 (online)

OBSAH

Venujeme

prof. Ing. Vladimírovi Kvasničkovi, DrSc.,

nestorovi umelej inteligencie
a kognitívnej vedy na Slovensku.

PREDHOVOR

Idea vydania zborníka o umelej inteligencii (UI) sa nezrodila vo vzduchoprázdne. Predchádzal jej teoretický interdisciplinárny seminár na tému *Umelá inteligencia: sociálne a etické problémy,* ktorý sa uskutočnil 24.mája 2023 v rámci Slovenskej sociologickej spoločnosti na pôde Sociologického ústavu SAV. Bolo to spoločné podujatie viacerých sekcií tejto spoločnosti, a to sekcie sociologickej teórie, sekcie sociológie kultúry, sekcie sociológie zdravotníctva a Východoslovenskej pobočky. K ich spolupráci došlo už po februári 2023, hoci už v novembri 2022 vedúca sekcie sociologickej teórie doc. D. Alieva prišla s návrhom organizácie spoločného seminára rôznych sekcií SSS na tému umelej inteligencie. Nebolo to teda ešte pod vplyvom „boomu" okolo UI, ktorý o pár mesiacov ovládol slovenské médiá v súvislosti so zverejnením jazykového modelu ChatGPT-3. Možno aj preto návrh doc. D. Alievy zo začiatku nenašiel jednoznačnú podporu na pôde SSS. Niektorí z vedúcich sekcii, ktorých oslovila v januári 2023, dokonca vyjadrili pochybnosti o aktuálnosti tejto témy pre slovenskú sociológiu.

Našťastie, nie všetci oslovení sa držali tohto názoru a naopak vrele podporili ideu organizácie spoločného teoretického seminára o umelej inteligencii a s ňou súvisiacich sociálnych, právnych a etických problémoch. Boli to vedúca sekcie sociológie kultúry doc. A. Kvasničková, vedúca sekcie sociológie zdravotníctva S. Capíková a vedúca Východoslovenskej pobočky G. Brutovská. Každá z nich sa stretla s fenoménom umelej inteligencii vo svojej vedeckej, pedagogickej a odbornej práci. Niektorí z nich sa zúčastnili konferencií na témy, súvisiace s automatizáciou, digitalizáciou a robotizáciou, dokonca mali podiel na ich organizácii. Niektorí z nich už aj publikovali k tejto problematike. Dalo sa predpokladať, že ich účasť už na príprave seminára by veľmi prispela k jeho realizácii. Veď ich prostredníctvom sa zabezpečoval

kontakt so sférami kultúry, zdravotníctva, sociálnej práce, sociálnej starostlivosti, a pod., kde si UI nachádzala najširšie uplatnenie. Poznatky, praktické skúsenosti a v značnej miere aj profesiové kontakty, ktoré získali v tejto oblasti, mali prispieť ku konkretizácii tematiky seminára a k jeho naplneniu aktuálnou problematikou, súvisiacou hlavne s aplikáciou UI. Táto úloha sa riešila výberom vhodných autoriek/autorov príspevkov, ktoré by mali odznieť na seminári. V tomto smere sa vedúci sekcií usilovali o to, aby obsadili seminár najlepšími odborníčkami a odborníkmi, čo sa im aj podarilo.

Veľmi sa o to pričinila doc. A. Kvasničková, ktorá mala na starosti kontakty s odborníkmi v sfére kognitívnej vedy a aplikovanej informatiky. Neváhala zaangažovať do toho aj svojho manžela prof. V. Kvasničku, ktorý ako zakladateľ kognitívnej vedy na Slovensku a uznávaný vysokoškolský pedagóg, pripravil mnoho významných odborníkov v tejto oblasti. Vďaka jeho profesiovým kontaktom a osobnej prosbe sa podarilo získať súhlas pre aktívnu účasť na seminári tak od prof. I. Farkaša, ako aj od prof. Ľ. Beňuškovej. Už samo osebe získať pre seminár týchto špičkových odborníkov z Centra pre kognitívnu vedu Katedry aplikovanej informatiky FMFI UK znamenalo napoly zabezpečiť jeho úspech u sociologického obecenstva. Išlo o to, aby ich príspevky, ktoré by odzneli v úvodnej časti seminára a ktoré by sa venovali výkladu základnej problematiky UI a perspektívam jej vývoja, hneď neodpudili eventuálnych návštevníkov príliš-nou náročnosťou a nezrozumiteľnosťou. Naopak, mali by im vysvetliť a dopovedať to, čo ešte ostáva za rámcom popularizačných výkladov o UI, ktorých sú plné dnešné média. Prípadne i zodpovedať niektoré „dráždivé" otázky ohľadom perspektív spolunažívania inteligencie ľudskej a tej umelej. A dá sa povedať, že práve tak sa tomu aj stalo na seminári 24. mája 2023.[1] Prednášky prof. I. Farkaša a prof. Ľ. Beňuškovej získali sympatie sociologického publika svojím zrozumiteľ-

[1] Podrobnejšia informácia o seminári v: Capíková, S., Tížik, M., *Umelá inteligencia: sociálne a etické problémy.* Správa o seminári Slovenskej sociologickej spoločnosti pri SAV. Bratislava, 24. 5. 2023. Sociologický časopis, č.4/2023, s. 475-479.

ným a pútavým výkladom náročných a obsiahlych tém. Autor a autorka pritom nezjednodušovali celkovú situáciu v súvislosti s ďalším rozvojom UI, ale otvorene poukazovali na isté rizika, ktorým spoločnosť bude musieť čeliť.

Obecenstvu najviac imponovalo aj to, že mali pred sebou autentických odborníkov na UI, ktorí im nič nezatajujú, a pritom ani nezjednodušujú, ako to robia pochybní vykladači tejto problematiky „z druhej ruky". Práve preto ich prednášky nielen pritiahli na seminár mimoriadne veľký počet návštevníkov, ale aj prebudili u nich živý záujem o ich obsah. Prejavilo sa to aj v ich otázkach k prednáškam, aj v ich krátkych diskusných príspevkoch. A hoci autormi otázok boli sociológovia, ich obsah svedčil o dostatočnej úrovni informovanosti o súčasnom stave UI. Z tohto hľadiska kontakt medzi odborníkmi na informatiku a sociologickým obecenstvom na seminári pripomínal nie tak poučovanie jednou stranou tej druhej, ako skôr dialóg partnerov, hľadajúcich vzájomné porozumenie ohľadom toho istého objektu, ktorým bola UI.

Treba povedať, že toto ovzdušie partnerského dialógu a konsenzu panovalo na seminári po celý čas jeho konania. A to aj počas druhej časti seminára, keď sa slova ujali sociológovia. Oni sa usilovali hlavne o to, aby predstavili na seminári aj sociologické koncepcie UI. No ich vystúpenia nemali konfrontačný ani polemický charakter. Naopak, aj v „sociologickej" časti seminára išlo o hľadanie spoločného jazyka a styčných bodov medzi sociológiou a informatikou. V tomto duchu bol ladený príspevok významného slovenského metodológa prof. J. Schenka, ktorý sa dlhodobo venuje multiagentovému modelovaniu (MAM) v sociológii. Práve MAM je, aj podľa autora, distribuovanou umelou inteligenciou. MAM sa stalo efektívnym nástrojom skúmania multikomponentových nelineárnych dynamických systémov, medzi ktoré patria aj sociálne systémy. Čo sa týka českého sociológa J. Mlynářa, ten vníma UI ako bytostný sociologický problém a sociálny fenomén, ktorý sa nedá uzavrieť vo vnútri stroja, a ktorý sa prejavuje v jazykovej interakcii ľudí. Na tento názor sociológa J. Mlynářa už reagoval otázkou informatik I. Farkaš a dostalo sa mu odpovede, ktorá ho zrejme uspokojila.

V istom zmysle sa nielen sociológovia učili od informatikov, ale aj informatici zrejme niečo získali od kontaktov so sociológmi. A nielen s nimi. So záujmom počúvali príspevky prevažne aplikačného zamerania, s ktorými prišli na rad už odborníčky z oblasti sociálnej práce, ošetrovateľstva a zdravotníctva. Tak napr. podrobné rozbory tých závratných zmien v charaktere práce, ktoré prináša so sebou UI do výrobnej sféry (G. Brutovská – B. Balogová), určite budú nápomocné aj pre informatikov. Podobne aj bohatý prehľad možnosti využitia techník UI v rámci zdravotníctva a ošetrovateľstva. (I. Ondriová – T. Fertaľová). Prierezovou témou pre všetky odbory, používajúce UI, sa stávajú s tým súvisiace etické a právne otázky (M. Kolesárová). Tieto otázky sa stali predmetom diskusií na seminári. Iba na niektoré otázky sa podarilo počas semináru dostať jednoznačnú odpoveď od prítomných informatikov. Niektoré otázky však ostali nezodpovedané. Možno práve preto na organizátorov seminára sa už po jeho ukončení obrátili niektorí záujemcovia z obecenstva s prosbou o pokračovanie s témou UI, ako aj s prosbou o vydanie zborníka zo seminára. A hoci sme to neplánovali, predsa sme sa rozhodli najprv opýtať autorov príspevkov, či si prajú toto vydanie a či sa ho zúčastnia. Dostali sme od nich jednoznačný súhlas s vydaním zborníka, ako aj s ich osobnou účasťou na zborníku.

Vyvstala však otázka financovania vydania zborníka. Aj tu nám znova pomohla doc. A. Kvasničková. Prihovorila sa v prospech vydania zborníka u prof. P. Sinčáka, ktorý je podpredsedom Správnej rady AIslovakIA, a poprosila ho o finančnú podporu jeho vydania. Súhlasil podporiť vydanie zborníka z prostriedkov AIslovakIA. Navrhol nám, že najprv to bude elektronické vydanie zborníka, až potom budeme uvažovať o printovom vydaní. Úprimne ďakujeme i prof. P. Sinčákovi za jeho pomoc a podporu, ako aj doc. A. Kvasničkovej za jej príhovor u prof. P. Sinčáka. Takisto úprimne ďakujeme prof. V. Kvasničkovi za prejavenú podporu pri organizácii nášho seminára a prajeme mu veľa zdravia.

V súvislosti so zborníkom sme sa rozhodli zachovať rovnaký názov ako mal seminár a ponechať si aj také isté usporiadanie príspevkov, čo bolo na seminári. Preto zborník reprodukuje štruktúru a obsah semi-

nára, interdisciplinárny charakter ktorého sa v plnej miere odrazil aj v terajšej skladbe príspevkov. Vyprofilovali sa dve disciplinárne definované skupiny príspevkov. Aj v zborníku prvé dva texty pochádzajú z pera informatika a neurovedkyne. Napriek tomu, že sa zaoberajú fundamentálnymi problémami umelej inteligencie, obaja skúsení autori ich napísali s osobitným zreteľom na špecifickú kategóriu čitateľov, ktorým sú predovšetkým určené: informačne bohato, jasne, zrozumiteľne a veľmi prístupne. Druhú skupinu tvoria príspevky, ktorých autormi sú prevažne sociológovia alebo reprezentanti veľmi blízkych disciplín.

Igor Farkaš svojím príspevkom pod názvom *Umelá inteligencia a spoločnosť: príležitosti a riziká* načrtáva základný rámec zložitej problematiky umelej inteligencie a jej možných vplyvov na spoločnosť. Stručne charakterizuje umelú inteligenciu a jej doterajší vývoj. Osobitnú pozornosť venuje neurónovým sieťam a problematike učenia v tzv. hlbokých neurónových sieťach. Sústreďuje sa predovšetkým na inteligentné četboty a ChatGPT, ktorý od začiatku roku 2023 vzbudil mimoriadnu pozornosť a zintenzívnil vášnivé polemické diskusie v tejto oblasti, dokonca až do polohy prípadnej kontrolovanej regulácie rozvoja umelej inteligencie vôbec. Autor sa orientuje najmä na praktické prístupy k umelej inteligencii, ktoré sa týkajú možností zlepšenia kvality ľudského života, a naznačuje aj potenciálne riziká, ktorými spoločnosť môže byť umelou inteligenciou ohrozená.

Problém vedomia má v tejto oblasti nepochybne fundamentálny význam, no doteraz ostáva otvorený. Prednostne sa mu venuje vo svojom príspevku *QuoVadis UI?* Ľubica Beňušková. Vychádza z dobre overeného predpokladu, že vedomie má emergentný a evolučný charakter. Zdôrazňuje kľúčové poznatky o vedomí z filozofie a neurovedy, špecifiká vo zvieracej ríši i ich dôsledky pre informatiku. Významným prínosom je výpočtový funkcionalizmus, t. j. zoznam indikačných charakteristík vedomia (ako subjektívneho prežívania), ktoré sú obsiahnuté v 6 špecifických teóriách a ktoré by mal v plnom rozsahu spĺňať každý vedomý systém bez ohľadu na to, či jeho prvky tvoria neuróny alebo mikročipy. V súvislosti s vedomím filozofia i veda stále ostávajú pred voľbou medzi nejakou verziou materializmu, ktorá je redukcionistická a popiera existenciu subjektívnych stavov vedomia,

a nejakou verziou dualizmu, ktorá ich síce uznáva, no znamenala by, že žijeme v dvoch svetoch – fyzikálnom a mentálnom – čo je značne vágne. Otvorená je aj otázka, či redukcionistická verzia už bola dopracovaná, alebo bude vyžadovať zavedenie nejakých mentálnych entít alebo formuláciu nového prírodného zákona, ktorý opíše prepojenie medzi fyzikálnym a mentálnym. Ľudská myseľ napokon isto dokáže pochopiť a vysvetliť aj samu seba.

Príspevok *Distribuovaná umelá inteligencia v sociológii: multiagentové modelovanie* predstavuje aj akési prepojenie medzi informatikou a sociológiou. Juraj Schenk najprv stručne ilustruje základné charakteristiky tejto špecifickej metodológie (najmä v porovnaní s tzv. sociológiou premenných veličín) a hlavné znaky virtuálnych sociálnych agentov. Zdôrazňuje tri skupiny inšpirácií, ktoré sociológia získava aplikáciou multiagentového modelovania osobitne pri skúmaní sociálnej dynamiky. Prvou je možnosť simulácie v sociologickom laboratóriu, ktorú sociológia doteraz (skoro) vôbec nemala, druhou je kontraintuitívnosť výsledkov (dokumentovaná Schellingovým modelom segregácie osídlenia a modelom kognitívnej deľby práce) a treťou sú možnosti prekonávania „tradičných" limitov najmä v súvislosti s budovaním a testovaním sociologických teórií.

Jakub Mlynář nazval svoj text *Umělá inteligence jako sociologický problém*. Bežné interpretácie umelej inteligencie, ktoré sa opierajú o kognitivistické analógie, nepovažuje za dostatočne adekvátne. Východisko vidí vo wittgensteinovskej filozofii a prístupoch, ktoré na ňu nadväzujú: myseľ a jej koreláty sú pozorovateľné a zvonka prístupné javy, ktoré sú zakotvené v sociálnej interakcii. Na tomto základe bude možné vybudovať alternatívnu, bytostne sociologickú interpretáciu umelej inteligencie ako sociálneho fenoménu, ktorý je v reálnom čase vyprodukovaný konaním členov spoločnosti: umelá inteligencia vyvstáva zo špecifickej usporiadanosti detailov situovaného konania. Naznačený prístup ilustruje príkladmi, ktoré pochádzajú z analýzy využitia umelej inteligencie v medicínskom prostredí a interakcie s jazykovým modelom ChatGPT.

Vplyv digitálnych technológii a umelej inteligencie na charakter práce je tematika, o ktorej píše dvojica autoriek Gizela Brutovská a Beáta

Balogová. Pozornosť sústreďujú najmä na dôsledky, ktoré – v období spoločnosti Priemyslu 4.0 – v charaktere práce vyvoláva robotizácia a digitalizácia. Autorky osobitne analyzujú tri problémové okruhy. Sú to: 1. globalizácia trhu a vplyv digitálnych technológií, 2. flexibilizácia práce a jej hlavné formy a 3. obmedzenia umelej inteligencie a robotov vo vzťahu k človeku a práci. V súvislosti s nástupom umelej inteligencie zdôrazňujú, že ohrozené sú len rutinné pracovné miesta. Predstava, že umelá inteligencia pripraví v blízkej budúcnosti o prácu kvalifikovaných zamestnancov, je úplne chybná, osobitne v oblastiach, kde je potrebná ľudská interakcia a emočná inteligencia. Navyše, cena digitálnych technológií, najmä umelej inteligencie je neúmerne vysoká.

V poradí posledný príspevok *AI v medicíne. Eticko-právno-sociálne výzvy a medicínske vzdelávanie* od Márie Kolesárovej a Anamarie Malešević naznačuje, že v oblasti zdravotníctva sa problematike umelej inteligencie venuje zvýšená pozornosť. V prvom bloku sa autorky zaoberajú sa umelou inteligenciou v medicíne, potrebou normatívneho rámca, zdôrazňujú požiadavku, aby umelá inteligencia najmä v tejto sfére bola zodpovedná a diskutujú etické, právne a sociálne problémy, ktoré sa s umelou inteligenciou v medicíne spájajú. Druhý blok súvisí s úlohou umelej inteligencie v medicínskom vzdelávaní. Autorky tu rekapitulujú výsledky prieskumu na troch lekárskych fakultách v SR, kde sa získavali údaje od domácich i zahraničných študentov. Predmetom záujmu boli postoje k umelej inteligencii a jej využívaniu v medicíne, etická dimenzia vzťahu medzi lekárom a pacientom a názory na právnu reguláciu umelej inteligencie v medicíne.

Bratislava október – november 2023

Dilbar Alieva
Juraj Schenk

UMELÁ INTELIGENCIA A SPOLOČNOSŤ: PRÍLEŽITOSTI A RIZIKÁ

Igor Farkaš

Abstrakt
Od svojho vzniku v polovici minulého storočia prešla umelá inteligencia rôznymi etapami až do súčasnosti, keď sa o nej začalo oveľa intenzívnejšie hovoriť, a to v súvislosti s mnohými jej čiastkovými úspechmi. Tie sa týkajú rôznych konkrétnych úloh, či už klasifikácie obrázkov, hrania logických hier alebo spracovania prirodzeného jazyka. Za týmito úspechmi sa skrývajú moderné algoritmy strojového učenia v hlbokých neurónových sieťach. V súvislosti s úspechmi umelej inteligencie vznikli o nej početné vášnivé spoločenské debaty od filozofických (môže raz nadobudnúť vedomie?) po praktické (ako nám môže pomôcť zlepšiť kvalitu života?). Aby sme si ozrejmili, s čím máme do činenia, v príspevku stručne predstavíme umelú inteligenciu, hlavné charakteristiky neurónových sietí, ich silné i slabé stránky. Potom zameriame pozornosť na četovací program ChatGPT, keďže táto nová jazyková technológia rýchlo vzbudila svetový záujem v oveľa väčšej miere ako doteraz čokoľvek iné v oblasti umelej inteligencie. Napokon spomenieme možnosti jej využitia pre spoločnosť ako aj potenciálne riziká.

1 ČO JE UMELÁ INTELIGENCIA

Ľudský druh Homo sapiens, čiže človek rozumný, si zakladá na tom, že inteligencia je jeho kľúčovou charakteristikou. Teda to ako vníma, rozmýšľa, rozhoduje sa, koná, plánuje a interaguje v sociálnom kontexte. Sumárne to nazývame kognitívnymi procesmi. Človek

nepochybne dokáže vďaka svojej inteligencii riešiť široké spektrum problémov, sám alebo v spolupráci, a preto bol hlavným aktérom kultúrnej evolúcie, vďaka ktorej sme sa dostali až do súčasnej modernej, exponenciálne zrýchľujúcej sa doby (Tomasello, 2001). Na druhej strane, vďaka výskumom v oblasti kognitívnej psychológie vieme, že človek podlieha širokému spektru kognitívnych skreslení, v dôsledku ktorých sa častokrát rozhoduje nerozumne, iracionálne, aj pod vplyvom emócií a na základe rôznych heuristík (Pinker, 2022). Takže človek je tvor inteligentný, ale nezriedka nenachádza optimálne riešenia problémov.

Umelá inteligencia je vedecká disciplína, ktorej cieľom je riešiť pomocou výpočtových nástrojov rôzne problémy podobným spôsobom ako človek alebo aj inak a lepšie, v ideálnom prípade optimálne. Inšpiráciou tu môže byť kognitívna veda, ktorá skúma ľudské myslenie, ale aj všeobecne samotná matematika, ktorá je univerzálnym nástrojom na opis sveta a dá sa využiť na hľadanie optimálnych ciest. V oboch prípadoch pracujeme s predpokladom, že inteligentné procesy možno vysvetliť mechanisticky, pomocou matematických konceptov a operácií s nimi, teda pomocou formalizácie prístupu k riešeniu problému.

Druhým zaujímavým aspektom pri tvorbe inteligentných systémov je to, či sa naň pozeráme zvonka alebo zvnútra. Pohľad zvonka spájame so správaním, pohľad zvnútra s myslením, teda skrytými mentálnymi procesmi, ktoré vedú k nejakému správaniu. Umelá inteligencia nie je jeden smer, jeden pohľad, ale je spektrom rôznych metód, ktoré sa navzájom odlišujú podľa toho, na aké podproblémy sa zameriavajú a ako pristupujú k ich riešeniu. Aj v prípade človeka ide o celé spektrum schopností, počnúc rozpoznávaním obrazu a spracovaním jazyka, cez priestorovú navigáciu a motorické úkony, až po abstraktné rozhodovanie a plánovanie. Na všetky využíva ten istý mozog a telo, no stále do detailov nevieme ako. V prípade človeka je pohľad do jeho mysle komplikovaný, na rozdiel od umelých systémov, ktoré konštruujeme.

Systémy s umelou inteligenciou môžeme rozdeliť aj z toho pohľadu, ako sú vtelené. Poznáme roboty i softboty. Oba typy botov môžu disponovať inteligenciou, avšak roboty či autonómne autá sú pre člo-

veka viac viditeľné. Robotické systémy priamo zasahujú do prostredia alebo interagujú s človekom. Softboty, kam patria aj četboty, sú inteligentné agenty skryté v počítači. Nemajú telo, preto ich vnímame viac abstraktne. Obe kategórie botov sa permanentne vyvíjajú.

V rámci rýchleho technologického vývoja v 21. storočí sme sa dostali do fázy, keď sa o umelej inteligencii začalo oveľa častejšie hovoriť, a to vďaka výsledkom, ktoré boli publikované. Preto sa zamyslíme nad tým, čo vlastne umelá inteligencia je, čo sú jej nové výdobytky, a čo to znamená pre spoločnosť, ktorej sa tieto novinky bytostne dotýkajú. Osobitnú pozornosť budeme venovať četovaciemu programu ChatGPT, ktorý bleskurýchlo oslovil pred rokom milióny ľudí po celom svete, viac ako akákoľvek iná technológia UI predtým.

1.1 Definície umelej inteligencie

Neexistuje jediná definícia umelej inteligencie, na ktorej by sa zhodla celá vedecká komunita. Populárna učebnica umelej inteligencie (Russell a Norvig, 2010) uvádza štyri možnosti ako definovať umelý inteligentný systém, v závislosti od dvoch vyššie spomínaných nezávislých aspektov (tabuľka 1).

Prvá dvojica definícií sa vzťahuje na človeka. Umelá inteligencia *koná ako človek* vtedy, ak jej správanie sa javí pozorovateľovi ako ľudské. Tento pohľad zaviedol v roku 1950 Alan Turing v podobe testu, ktorý nesie jeho meno, a ktorý sa realizuje cez textovú obrazovku v prirodzenom jazyku. Pozorovateľné správanie generované četovacím programom (četbotom) alebo človekom je takto redukované na jazyk, pozorovateľ nevidí, čo alebo kto je v pozadí. Takýmto testom, ktorý má aj svoje nedostatky, donedávna neprešiel žiaden umelý systém (na rozdiel od četbota ChatGPT, o ktorom bude reč neskôr).

Druhý pohľad – inteligentný systém *myslí ako človek* – bol iniciovaný kognitívnou psychológiou, ktorá začala empiricky skúmať mentálne procesy. Teoreticky, keby sme vedeli takéto procesy spoľahlivo identifikovať, formalizovať a implementovať ich v umelom systéme, tak ten by spĺňal túto definíciu inteligencie. Tento prístup je však veľ-

mi obtiažny kvôli tomu, že mentálne procesy u človeka vieme skúmať len nepriamo, pretože sa chápu abstraktnejšie ako neurálne aktivity v mozgu. Vyššie spomínaný Turingov test mal za cieľ zistiť, či „stroj dokáže myslieť", avšak cez jeho vonkajšie prejavy (správanie).

Druhá dvojica definícií chápe racionalitu (optimálne riešenie) ako kľúčový predpoklad inteligencie. V tomto prípade to znamená, že ak vieme merať, ktoré prvky správania (akcie) sú v danej situácii optimálne, tak podľa tohto vieme usúdiť, či daný systém UI (agent) *koná racionálne*. Toto je možné v prípade konkrétnych úloh v jasne definovaných prostrediach (napr. hranie hier), nie však v reálnom komplexnom svete.

Napokon, pohľad, že UI *myslí racionálne,* je založený na pravidlách formálnej logiky a racionálnom usudzovaní, ktoré je opísateľné exaktným jazykom. Tu je však hlavným problémom to, že zďaleka nie všetky problémy (napr. motorický pohyb) sa dajú dobre riešiť pomocou pravidiel (typu „ak sa leskne a je veľmi tvrdý, potom diamant").

Tabuľka 1
Príklady definícií umelej inteligencie rozdelené do štyroch kategórií.

Správa sa ako človek „Umenie vytvárať stroje, ktoré vykonávajú funkcie vyžadujúce inteligenciu, keď ich vykonávajú ľudia." „Štúdia o tom, ako prinútiť počítače robiť veci, v ktorých sú v súčasnosti ľudia lepší."	Správa sa racionálne „Výpočtová inteligencia je štúdium dizajnu inteligentných agentov." „UI sa zaoberá inteligentným správaním v artefaktoch."
Myslí ako človek „Vzrušujúce nové úsilie prinútiť počítače myslieť ... stroje s mysľou v plnom a doslovnom zmysle." „[Automatizácia] činností, ktoré spájame s ľudským myslením, činnosťami, ako je rozhodovanie, riešenie problémov, učenie..."	Myslí racionálne „štúdium mentálnych schopností pomocou výpočtových modelov." „štúdium výpočtov, ktoré umožňujú vnímať, uvažovať a konať."

1.2 Stručná história UI

Pre pochopenie súčasnosti je dobré poznať minulosť. Umelá inteligencia ako vedecká disciplína sa zrodila v polovici minulého storočia (Russell a Norvig, 2010), v čase rozkvetu kybernetiky, ktorej zakladateľom bol brilantný americký matematik a filozof Norbert Wiener. Jeho celoživotným cieľom bolo zjednotiť matematickú teóriu, elektroniku a automatizáciu do „teórie riadenia a komunikácie u zvierat a strojov". Druhým vplyvným aktérom toho obdobia, ktorého hodno spomenúť je maďarsko-americký matematik, fyzik, inžinier a počítačový vedec v jednej osobe John von Neumann, vynálezca digitálneho počítača, ktorý znamenal nástup tretej priemyselnej revolúcie. Do tretice, už vyššie spomenutý brilantný britský matematik a logik Alan Turing, ktorý sa spoznal s von Neumannom už v polovici 30-tych rokov, nezávisle tiež vytvoril princípy elektronického počítača a zaviedol pojem Turingov stroj, ktorý je konceptuálnym základom pre teoretickú informatiku. Preslávil sa však aj tým, že počas druhej svetovej vojny skonštruoval zariadenie, ktoré dokázalo odhaliť šifrovací kód (Enigma) používaný armádou nacistického Nemecka, čo prispelo k víťazstvu Spojencov. Všetky tu spomínané vynálezy a teoretické poznatky prispeli k zrodu umelej inteligencie.

Jedným zo zakladateľov UI je americký počítačový a kognitívny vedec John McCarthy z MIT, ktorý UI vnímal ako „výpočtovú časť schopností dosahovať ciele vo svete" (McCarthy, 1999). Druhý zakladateľ, Marvin Minsky z Carnegie-Mellon University v USA definoval UI ako „konštrukciu počítačových programov, ktoré sa zaoberajú úlohami, ktoré v súčasnosti uspokojivejšie vykonávajú ľudské bytosti." Obaja skvelí vedci sa zúčastnili v roku 1956 na prelomovom workshope v rámci konferencie na Dartmouth College v USA, kde položili základy vzniku a rozvoja umelej inteligencie. Vývoj UI nebol priamočiary a lineárny, ale ako to už býva, zaznamenal svoje silnejšie a slabšie etapy (Russell a Norvig, 2010). Pre účely tohto príspevku postačí, ak poukážeme na rozdiely medzi dvoma hlavnými, kvalitatívne odlišnými kategóriami prístupov v umelej inteligencii – symbolizmom a konekcionizmom (Šefránek, Takáč, Farkaš, 2008; Farkaš, 2011).

1.3 Symbolové systémy a konekcionizmus

Prvé dekády rozvoja UI sa niesli v znamení tzv. symbolovej paradigmy, ktorá je založená na princípoch diskrétnej, symbolovej reprezentácie znalostí (tzv. ontológie), logiky a pravidiel. V tých časoch dominovalo presvedčenie, že prístup „zhora" umožní priamo naprogramovať UI na riešenie konkrétnych problémov. Takto napríklad ešte v bývalom Československu vznikol expertný systém na medicínsku diagnostiku (Popper, 1986). Ukázalo sa však, že v zložitejších prípadoch je táto cesta veľmi problematická, napríklad preto, že počet potrebných pravidiel bude neúnosne vysoký, alebo že nebudeme vedieť, aké všetky pravidlá budú potrebné. Taktiež, nie všetko sa dá dobre opísať pomocou (hrubých) pravidiel, a často existuje veľa výnimiek. Toto samozrejme závisí od typu problému. Napríklad rozhodovací algoritmus na udeľovanie víz môže dobre fungovať, no presná detekcia zhubného nádoru na tomografickej snímke pomocou pravidiel je problematická. Preto začala byť zaujímavá myšlienka nesnažiť sa nadizajnovať riešenie priamo, ale nechať systém naučiť sa ho pomocou tréningu.

Takýto alternatívny prístup je základom konekcionistických systémov, teda umelých neurónových sietí, ktoré sa vedia učiť. Neurónové siete sa začali skúmať neskôr ako symbolové systémy a ich vývoj tiež nebol kontinuálny, a to kvôli vtedajším problémom, ktoré čakali na svoje vyriešenie (napr. ako správne realizovať tréning siete). Umelá neurónová sieť pozostáva z množiny navzájom poprepájaných (preto konekcionizmus) jednoduchých prvkov – umelých neurónov, ktoré navzájom komunikujú cez váhované spojenia, a tie predstavujú dlhodobú pamäť systému. Sila umelej neurónovej siete spočíva práve v interakciách medzi neurónmi a v zmenách sily spojení, ktoré sprevádzajú proces učenia (Kvasnička a kol., 1997; Sinčák a Andrejková, 1986). Znalosť v neurónovej sieti existuje nie v podobe symbolov, ale v podobe reálnych čísel (vyjadrujúcich sily spojení), preto sú neurónové siete sybsymbolové. Je tu zrejmá biologická inšpirácia prístupu. Paralelné spracovanie informácie a distribuovanosť (roztrúsenosť) znalostí predstavujú základný konceptuálny rozdiel v porovnaní so

symbolovými systémami (Farkaš, 2011). Umelé neurónové siete sú teda prístupom založenom na strojovom učení, ktoré v ostatnej dekáde hrá v oblasti UI primárnu úlohu.

1.4 Strojové učenie v hlbokých neurónových sieťach

Hlboké učenie v neurónových sieťach je v súčasnosti dominujúcim prístupom k úspešnému riešeniu pomerne širokého spektra problémov, vďaka čomu sa UI teší vysokej popularite v rôznych aplikačných oblastiach (Schmidhuber, 2015). Patrí sem napríklad spracovanie obrazových dát (klasifikácia do tried), úlohy v prirodzenom jazyku či sekvenčné rozhodovacie úlohy (napr. hranie hier proti ľudským súperom). Presnosť týchto výpočtových modelov v mnohých prípadoch dosahuje úroveň človeka, a niekedy ho aj prekonáva (Mnih a spol., 2015). Učenie na príkladoch sa javí ako najlepší spôsob, ako dosiahnuť zložité správanie, ktoré formálne predstavuje matematické zobrazenie (transformáciu) vstupov na výstupy, napríklad obrázku na predikovanú kategóriu alebo časti vety na jej pokračovanie. Toto zobrazenie, implementované hlbokou neurónovou sieťou sa často označuje ako „čierna skrinka".

Ukázalo sa, že hlboká neurónová sieť, ktorá má viacero vrstiev medzi vstupmi a výstupmi, funguje lepšie ako plytká sieť s malým počtom vrstiev. Toto by nemalo byť prekvapujúce, pretože hlboká sieť má viac váh medzi neurónmi, a teda parametrov, ktoré sa musia natrénovať. V tejto súvislosti je zaujímavé poznamenať, že bolo preukázané, že hĺbka sa nedá efektívne nahradiť šírkou, teda tak že by sme dosiahli rovnaké množstvo parametrov v sieti s malým počtom skrytých vrstiev ale s veľkým počtom neurónov. Z matematického hľadiska hĺbka umožňuje efektívnejšie kombinovať znalosti do komplexnejších foriem. Napríklad, úloha rozpoznať čo je na farebnom obrázku a opísať to slovne je pre človeka obyčajne jednoduchá, pre neurónovú sieť je to však zložitá transformácia od pixelov (subsymbolová obrazová informácia na najnižšej možnej úrovni) po prirodzený jazyk (vysoká úroveň symbolovej sekvencie). Neurónová sieť sa to však dokáže naučiť, ak je dostatočne hlboká.

V neurónových sieťach sa využívajú tri paradigmy učenia: kontrolované (s učiteľom), nekontrolované (bez učiteľa) a učenie posilňovaním. Hlboké neurónové siete učiace sa s učiteľom využívajú čisto empirický prístup, pri ktorom sa sieť učí úlohy priamo z pôvodných anotovaných dát (tzv. end-to-end), teda párov vstup-výstup, a nerieši sa explicitne extrakcia príznakov. Sieť pri kontrolovanom učení v každom kroku porovná svoju predikciu so správnym výstupom, a podľa vzniknutej chyby si svoje parametre (váhy spojení medzi neurónmi) zmení trochu tak, aby sa nabudúce viac priblížila k správnej odpovedi (zmeny nemôžu byť príliš rýchle, aby sa systém nedestabilizoval).

Využíva sa aj *učenie posilňovaním* napr. na spomínané sekvenčné úlohy s rozhodovaním sa v čase, ale aj vo veľkých jazykových modeloch (ktoré sú napríklad základom četbotov ako ChatGPT) na dolaďovanie modelu pomocou spätnej väzby od užívateľov. Pri tomto type učenia sa sieť snaží získavať čo najvyššiu odmenu z prostredia (alebo užívateľa), preto si mení parametre tak, aby sa jej to viac a viac darilo.

Nech sa sieť učí akokoľvek, ultimátnym cieľom prístupu je dosiahnuť, aby svoje nadobudnuté poznatky sieť uplatnila aj v nových situáciách. Nemôžeme sieti ukázať obrázky všetkých psov na svete, aby ich vedela rozlišovať. Podstatou učenia je *extrahovať všeobecné črty* z konečnej množiny trénovacích obrázkov, a na základe nich správne klasifikovať (generalizovať) nové obrázky v testovacej fáze.

Úspech neurónových sietí neprišiel automaticky, ale bol umožnený dvoma hlavnými faktormi: prvým je technologický pokrok v počítačovom hardvéri (rýchlosť procesorov, paralelizácia výpočtov na grafických kartách, a veľkosť pamäte), vďaka ktorému je možné aj veľké modely natrénovať v „dohľadnom čase“ (t. j. týždne a mesiace), samozrejme v závislosti of hardvéru. Druhým dôležitým faktorom je dostatok digitálnych dát na trénovanie modelov, čo umožňuje zvyšovať ich presnosť. Množstvo trénovacích datasetov každým rokom rastie, pričom súťaživosť pri prekonávaní rekordov na benchmarkoch (verejne dostupných datasetoch) je hnacou silou zlepšovania algoritmov trénovania.

Výraznou charakteristikou súčasných neurónových sietí je, že už dokážu aj *generovať dáta*, nielen ich rozpoznávať, asociovať alebo klasifikovať, čo dokážu pasívne diskriminatívne modely. Generatívne

modely sú novou kvalitou, dokážu aktívne vytvárať nové dáta, a to v rôznych modalitách, ako obrázky, text, hudbu alebo video. Je to možné preto, lebo sú tak trénované, vytvoria si model obsahujúci charakteristiky existujúcich dát, na základe ktorých potom zovšeobecňujú. Vznikajú tak realisticky vyzerajúce tváre neexistujúcich ľudí, originálne texty alebo zaujímavo znejúce hudobné kúsky. Človek môže posúdiť kvalitu týchto výsledkov.

Takáto vlastnosť generatívnych modelov pochopiteľne vyvoláva aj obavy, pretože sme sa už dostali do fázy, keď nevieme čo na internete je pôvodné a čo syntetizované. Doteraz sme ako ľudstvo žili vo svete, ktorý sme pretvárali, aj tvorbou priamo vytváraných artefaktov, a teraz tieto artefakty dostávajú aj digitálnu podobu. Význam digitálneho sveta narastá, bez neho si už nevieme ani život predstaviť, no vzhľadom na to, že UI prispieva k zaplavovaniu digitálneho sveta vlastnými dátami, vzniká tu problém, čo je realita. Už teraz prebieha ťažký boj s dezinformáciami na internete, ktoré vyrábajú ľudia, teraz polienko do ohňa prikladá aj UI. O perspektívach umelej inteligencie si ešte povieme, aj preto, že súčasná UI má aj svoje nedostatky.

1.5 Nedostatky hlbokých modelov

Napriek svojim úspechom majú mnohé súčasné systémy s UI založené na strojovom učení aj rôzne nedostatky, a to konkrétne absentujúca robustnosť, nízka transparentnosť a zaujatosť (voči nedostatočne zastúpeným triedam dát v klasifikačných úlohách). Prvé dva nedostatky si ozrejmíme bližšie.

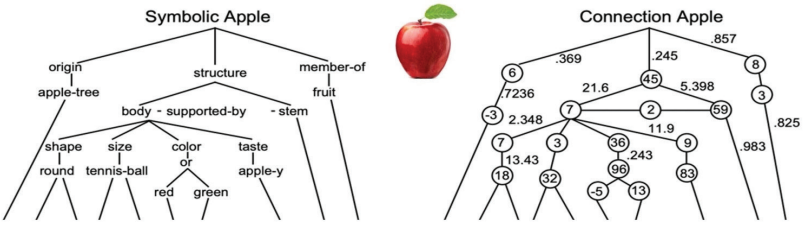
Absentujúca robustnosť znamená, že neurónové siete sú zraniteľné voči nepostrehnuteľným, tzv. adverzariálnym útokom. Absencia robustnosti natrénovaných neurónových sietí je najnovšie identifikovaný problém, ktorý bráni v nasadzovaní týchto modelov do rôznych kľúčových aplikácií. Tento problém prakticky znamená, že sieť slabo generalizuje svoje znalosti a dá sa ľahko oklamať. Samozrejme, nie hocijako, ale oveľa triviálnejšie, než človek. Geniálna myšlienka autorov Szegedy a spol. (2014) spočívala v návrhu takých špeciálnych vstu-

pov pre úspešne natrénovanú sieť, pre ktoré dáva úplne zlé predikcie, častokrát s vysokou mierou presvedčenia (konfidencie). Najbežnejším príkladom je klasifikácia obrázkov do tried. Natrénovaná sieť s vysokou presnosťou síce predikuje správne triedy na testovacích dátach (správne generalizuje), no napriek tomu sa dá ľahko oklamať obrázkami, ktoré boli len málo, no veľmi špecificky, pozmenené. Skúmanie robustnosti neurónových sietí patrí medzi aktívne oblasti výskumu (napr. Bečková, Pócoš, Farkaš, 2020).

Nízka transparentnosť neurónových sietí priamo vyplýva z ich architektúry a z reprezentácie znalostí (pomocou reálnych čísel), ktoré sú ukryté vo váhach medzi neurónmi a v aktivitách neurónov. Skrátka na prvý pohľad nevieme, čo natrénovaná neurónová sieť robí, lebo znalosti v nej sú roztrúsené. Avšak vzhľadom na úspešnosť týchto modelov je dôležité hľadať spôsoby, ako neurónovým sieťam lepšie porozumieť a zvýšiť tak ich dôveryhodnosť. Vysvetliteľná umelá inteligencia sa stala dôležitou vetvou výskumu, zameranou na pochopenie rôznych metód umelej inteligencie (Barredo-Arrieta a spol., 2020). Súčasťou tejto agendy je aj vysvetlenie toho, prečo neurónová sieť dáva na výstupe to, čo dáva (Montavon, Samek, Müller, 2018) alebo čo sa deje s dátami pri prechode zo vstupu na výstup (Pócoš, Bečková, Farkaš, 2022). Vysvetlenia sú dôležité pre rôzne cieľové skupiny, či už expertov, užívateľov alebo pacientov, s čím súvisia aj rôzne úrovne vysvetlenia (expert rozumie aj matematickým vzorcom, zatiaľ čo bežný človek uprednostní vysvetlenie v prirodzenom jazyku alebo pomocou obrázkov). Napríklad, lekár by pacientovi vedel vysvetliť pomocou takýchto metód, prečo si UI „myslí“, že nejde o zhubný nádor.

1.6 Transparentná a dôveryhodná UI

Vysvetliteľnosť v umelej inteligencii je dôležitou požiadavkou, uľahčujúcou jej širšie nasadzovanie v praxi, vrátane kritických aplikácií. S tým aj transparentnosť. Tú si vieme zjednodušene vysvetliť pomocou rozdielu medzi symbolovým systémom a neurónovou sieťou pri riešení konkrétnej úlohy, čo ilustruje obrázok 1.



Obr. 1
Ilustrácia klasifikátora na báze symbolovej umelej inteligencie (vľavo) a umelej neurónovej siete (vpravo). Symbolový systém postavený na formálnom jazyku je človeku zrozumiteľnejší než neurónová sieť, ktorá pracuje s číslami.
Podľa Bengio, Courville a Vincent (2013).

Oba prístupy dokážu rozpoznať, aký objekt je na obrázku. Symbolový systém dospeje k záveru sémantickou analýzou objektu, jeho častí, pričom jednotlivé kroky analýzy sú dobre vysvetliteľné. Na druhej strane, aj neurónová sieť vie dospieť k správnemu záveru, no keďže pracuje len s reálnymi číslami, jej jednotlivé kroky analýzy (medzivýsledky) nie sú ľahko vysvetliteľné, a teda málo transparentné. Avšak v prípade mnohých úloh sa ukázalo, že cesta práve cez učiace sa neurónové siete je najschodnejšia.

Cieľom vývoja UI je dosiahnuť, aby systémy UI boli čo možno najviac transparentné, vysvetliteľné, spoľahlivé, bezpečné (týka sa najmä robotických systémov) a tým pádom dôveryhodné (Farkaš, 2023). Súčasné systémy UI nie sú veľmi dôveryhodné, preto ich všeobecné využitie má svoje obmedzenia. Rozdiely existujú aj v prípade vtelených systémov UI (napr. robotov), ktoré tým, že majú telo, majú potenciál získať dôveru človeka v konkrétnejšej podobe než abstraktnejšie softboty (Fortunati a spol., 2015).

2 INTELIGENTNÉ ČETBOTY

Aj inteligentné četboty majú svoje obmedzenia, no napriek tomu ako horúca novinka vniesli do UI novú vlnu eufórie. Najznámej-

ším príkladom tejto takzvanej konverzačnej umelej inteligencie je ChatGPT, ktorý koncom novembra 2022 dala spoločnosť OpenAI k dispozícii verejnosti (verziu 3.5). Táto technológia chytila za srdce verejnosť viac ako akákoľvek iná technológia, čo ilustruje fakt, že v priebehu púhych päť dní získala milión registrovaných užívateľov a v januári nasledujúceho roka ich bolo už sto miliónov. Ľudia na celom svete začali produkt interaktívne „oťukávať", pretože sa s ním dá komunikovať v prirodzenom jazyku, a to nielen v angličtine. A netreba k tomu žiadne špeciálne vzdelanie či schopnosti, stačí mať počítač a prístup na internet.

ChatGPT je príkladom četbota, ktorý dokáže s človekom komunikovať cez textový terminál. Prvé četboty existovali už pred viac ako pol storočím (napr. ELIZA zo 60-tych rokov minulého storočia), v porovnaní s ChatGPT však boli veľmi „hlúpe" a človeku bolo jasné, že sa baví s počítačovým programom, ktorý nedisponuje inteligenciou. V prípade ChatGPT má však užívateľ celkom iný dojem, pretože ten je založený na veľkom jazykovom modeli a vyzerá preto oveľa inteligentnejšie.

Veľký jazykový model (ďalej VJM), ktorý stojí v pozadí, si môžeme predstaviť ako skrinku, ktorá má svoje vstupy a svoje výstupy (Zhao a spol., 2023). Táto skrinka teda dokáže na dané vstupy, napríklad časť vety, adekvátne odpovedať svojím výstupom, napr. pokračovaním vety. Tieto odpovede pritom netreba dizajnovať, ale dajú sa naučiť, pretože existujú k tomu dáta, samotné texty na internete. Skrinka má podobu obrovskej umelej neurónovej siete, učiacej sa štatistické zákonitosti v jazyku, z ktorých si vytvorí pravdepodobnostný prediktívny model.

Ak sa VJM učí predikovať ďalšie slovo vo vete, svoju chybnú predikciu vie korigovať na základe správnej odpovede. Napríklad, model číta sekvenciu „Peter sa zamiloval do…" a má predpovedať ďalšie slovo, ktorým je v danej vete „Evy" (v inom texte to môže byť napríklad „umenia"). VJM sa učí z vlastných chýb, podobne ako dieťa sa učí rozprávať, a keď dispozícii je rodič alebo učiteľ, môže ho opraviť.

2.1 A čo znamená GPT?

GPT je skratka pre generatívny predtrénovaný transformér. Toto slovné spojenie bežnému čitateľovi asi veľa nepovie, lebo každé z týchto slov si vyžaduje vysvetlenie. Transformér je najnovší typ architektúry VJM na báze veľkej neurónovej siete, ktorá využíva psychologicky inšpirovaný koncept pozornosti (Farkaš a spol., 2022), vďaka ktorému sa VJM naučí priraďovať rôznu mieru dôležitosti jednotlivým slovám pri predikcii očakávaného výstupu. V literatúre sa predtým používali aj iné úspešné modely na spracovanie prirodzeného jazyka, konkrétne rekurentné neurónové siete, avšak transformery sa ukázali ako úspešnejší nástroj, ktorý si vie lepšie zapamätať širší kontext a dosahovať tak vyššiu presnosť predikcie.

Predtrénovanie znamená, že daný VJM sa naučil z obrovského množstva textových dát (knihy, noviny, články na internete). Vo všeobecnosti sa dá povedať, že čím viac digitálnych dát použitých na trénovanie, tým presnejší model. Angličtina je na tom najlepšie z hľadiska početnosti textov. Po natrénovaní sa takýto VJM môže dotrénovať na konkrétne jazykové úlohy ako napríklad na detekciu sentimentu v texte, alebo rozpoznávanie pomenovaných entít (vlastných mien). Četovanie je tiež jednou z aplikácií, evidentne najzaujímavejšou.

Generatívnosť, ktorú sme už spomínali vyššie, je veľmi silná vlastnosť modelu, vďaka čomu dokáže generovať aj celé kusy textu. ChatGPT k tomu potrebuje nejaký podnet (nazývaný prompt) od užívateľa, čo môže byť jedna otázka, alebo aj kus textu, ktorý chce užívateľ dať povedzme parafrázovať.

2.2 ChatGPT prešiel Turingovým testom inteligencie

V polovici minulého storočia, v čase zrodu umelej inteligencie, brilantný britský matematik Alan Turing navrhol spôsob ako testovať inteligenciu stroja. Postavil sa k tomu z behaviorálneho hľadiska tak, že stroj (algoritmus) bude komunikovať s užívateľom v anglickom

jazyku cez obrazovku počítača (aby sme zámerne vylúčili akýkoľvek iný informačný kanál). Ak užívateľ po nejakom čase nadobudne dojem, že komunikuje s človekom, tak o testovanom stroji budeme môcť povedať, že je inteligentný. (Presnejší dizajn experimentu bol trochu zložitejší, ale podstata je rovnaká.)

V súčasnosti môžeme konštatovať, že ChatGPT uspel z pohľadu Turingovho testu. Napríklad preto, že viacero odborníkov keď ohodnocovalo kvalitu esejí napísaných ChatGPT a študentami, nevedeli ich pôvod rozlíšiť. Čo to však hovorí o porozumení textu technológiou ChatGPT?

2.3 Rozumie ChatGPT textom ktoré generuje?

ChatGPT generuje odpovede, faktom však je, že netuší, aké sú dobré, lebo sa mu vieme pozrieť „pod kapotu" (Takáč, 2023). Odpovede v drvivej väčšine prípadov dávajú zmysel, a čo je tiež dôležité, sú gramaticky správne. Takže s gramatikou tu problém nie, tú sa podarilo vďaka trénovaniu na veľkých dátach zvládnuť. Problémom je sémantika. Tú sa snažili tvorcovia programu vylepšiť aspoň tak, že pri dotrénovavaní ChatGPT využili aj učenie posilňovaním, v rámci ktorého kvalita odpovedí četovacieho programu bola ohodnocovaná viacerými ľuďmi (na danej stupnici). Táto forma spätnej väzby bola potom využitá na zlepšenie generatívnych vlastností modelu.

Napriek tejto procedúre, ktoré priniesla zlepšenie, stále ostáva problém reprezentácie významov v modeli. To ako VJM nadobúda znalosť o jazyku sa dá predstaviť tak, že by sme sa snažili naučiť jazyk z množstva čínskych textov, pričom čínštine nerozumieme. Naučili by sme sa, ako sa slová vyskytujú v jednotlivých kontextoch, ale samotným slovám by sme stále nerozumeli, pretože by to boli neznáme symboly, ktorých význam nie je ukotvený v skúsenosti so svetom. Človek sa učí jazyk práve v spájaní slov a fráz so skúsenosťou v okolitom svete. Slová sú symboly, ktoré asociujeme s ich významami.

Napriek tomuto fenoménu, dobre známemu v kognitívnej vede (interdisciplinárna oblasť zameraná na skúmanie ľudskej mysle),

v odbornej komunite umelej inteligencie prebiehajú vášnivé diskusie o tom, či ChatGPT rozumie. Zástancovia pozície, že áno, argumentujú čisto behaviorálne, Turingovým testom. Avšak ChatGPT si neraz aj vymýšľa, „halucinuje", tým že nemá kontakt s realitou. Taktiež zlyháva na jednoduchých logických úlohách (Biever, 2023). Toto vážne nahlodáva dôveryhodnosť takejto technológie.

O týchto nedostatkoch autori samozrejme vedia, a intenzívne pracujú na ich odstraňovaní. Súčasná, spoplatnená verzia (ChatGPT+) postavená na GPT-4 (OpenAI, 2023) je v porovnaní s pôvodnou, voľne dostupnou verziou založenej na GPT-3.5 výrazným skokom vpred. Snaha o napredovanie je poháňaná aj existujúcou konkurenciou, pretože už aj spoločnosť Google uviedla na trh četovací program Bard, ktorého základom je tiež VJM, a ktorý od júla 2023 je dostupný aj pre slovenčinu.

3 PRÍLEŽITOSTI A RIZIKÁ VYUŽITIA UI PRE SPOLOČNOSŤ

O umelej inteligencii a jej spoločenskom vplyve môžeme hovoriť v širšom kontexte. Začnime konkrétnejšie o všeobecne použiteľných četbotoch, pretože tie zaujali súčasnú spoločnosť najviac.

3.1 Četboty v budúcnosti

Četboty majú obrovský potenciál zmeniť spôsob, akým budeme komunikovať s technológiami a strojmi. Súčasne vďaka svojej generatívnej vlastnosti nám už ponúkajú svoje hotové „produkty". Mnohí študenti neodolávajú pokušeniu a aktívne využívajú ChatGPT na písanie esejí či záverečných prác. Tým pádom narážame na autorský problém nového typu, keďže tu nejde o klasické plagiátorstvo. Predložený text však tak či onak nie je vlastný. No toto nie je bežnými antiplagiátorskými programami zistiteľné.

Niektoré krajiny či inštitúcie sa rozhodli prístup k ChatGPT zablokovať, iné sa k tomu postavili pozitívne s návrhom ho rozumne

využívať pri práci či štúdiu. Bude však treba zmeniť prístup pri hodnotení, pretože odovzdávané textové produkty môžu mať „četbotovský pôvod". Možno sa zvýši dôraz na ústne skúšanie, kde sa testujú vlastné vedomosti. Je však evidentné, že ChatGPT sa ukazuje aspoň ako dobrý pomocník pri hľadaní odpovedí na široké spektrum otázok, a je efektívnejší ako klasické googlenie, pretože ponúka konkrétne štruktúrované odpovede aj väčšieho rozsahu. Avšak, zatiaľ bez záruky správnosti a pravdivosti.

Vývoj tejto technológie je veľmi ťažké predpovedať, dá sa však očakávať, že četboty sa budú výrazne zlepšovať. V širšom kontexte treba povedať, že súčasný vývoj umelej inteligencie je veľmi rýchly. Až tak, že spoločnosť na to nie je pripravená. Medzinárodný Inštitút Future of Life prišiel s iniciatívou otvoreného listu, v ktorom apeluje na medzinárodné spoločnosti a laboratóriá pracujúce v oblasti UI, aby minimálne na 6 mesiacov pozastavili trénovanie modelov lepších ako ChatGPT-4. List už podpísalo viac ako 33 tisíc ľudí. Nakoľko to pomôže, nevieme, pretože ide nielen o svetovú prestíž, ale koniec koncov o peniaze.

3.2 Umelá inteligencia všeobecne

Napriek svojim obmedzeniam ktoré sme spomínali vyššie, sa umelá inteligencia už integrovala do rôznych aspektov spoločnosti, pretvára priemyselné odvetvia, ovplyvňuje rozhodovacie procesy a bezprecedentným spôsobom ovplyvňuje náš každodenný život. Mnohé aplikácie UI v bežných technológiách (napríklad modul detekcie tváre v zabudovanej kamere v mobile) si ani neuvedomujeme a týkajú sa celej populácie, iné zase rôznych odvetví národného hospodárstva ako sú priemysel, poľnohospodárstvo, doprava, obchod, verejná správa, školstvo či obrana.

Debaty o umelej inteligencii v podstate existujú na dvoch hlavných úrovniach, nazvime ich praktická a filozofická. Praktické debaty sledujú pragmatický cieľ, ako sa postaviť k realite konštruktívne a využiť UI pre blaho spoločnosti už v súčasnosti. Na druhej strane, filozofic-

ké debaty sa snažia predpovedať, či raz UI nadobudne vedomie (viď príspevok Ľubice Beňuškovej), alebo či potenciálne pohltí ľudskú civilizáciu, lebo „zistí", že to pre ňu bude výhodné. Existuje pomerne vplyvný názor, že vďaka akcelerácii UI raz dosiahne úroveň superinteligencie (tzv. technologickej singularity), ktorá prevýši ľudskú inteligenciu a zásadne zmení každodenný život ľudí, pretože superinteligencia bude „všade". Napríklad známy vizionár Ray Kurzweil (2005) predpovedá, že singularita nastane v roku 2045.

My sa v tomto príspevku zameriavame na praktickú úroveň. Rozvinuté krajiny si uvedomujú dôležitosť UI, a preto sa snažia podľa svojich možností do nej investovať. Väčšina známych produktov UI pochádza z USA, ktoré sú lídrom, ale aj Európska únia sa tiež snaží konať. Vo februári 2020 vydala Európska komisia „Bielu knihu o umelej inteligencii – európsky prístup k dokonalosti a dôvere", ktorej cieľom je poskytnúť definíciu UI, zdôrazniť jej výhody a technologický pokrok v rôznych oblastiach, ale aj identifikovať jej potenciálne riziká. Na základe európskej stratégie pre UI predstavenej v apríli 2018 je súčasná Biela kniha komplexným dokumentom analyzujúcim silné a slabé stránky a príležitosti Európy na globálnom trhu umelej inteligencie. Európska únia schválila prvý zákon o umelej inteligencii (AI Act, 2023), ktorý identifikuje a opisuje rôzne miery rizika, od minimálneho po neakceptovateľné, ktoré sa budú využívať pri regulácii UI.

V rokoch 2018-19 boli jednotlivé členské štáty Európskou komisiou vyzvané, aby pripravili národné koncepcie rozvoja UI, a tak kolektív odborníkov vypracoval pre Úrad podpredsedu vlády SR pre investície a informatizáciu analýzu a návrh možností výskumu, vývoja a aplikácie UI na Slovensku. Vznikol komplexný dokument, ktorý obsahoval aj manuál pre firmy na zavedenie UI do praxe (Bieliková a kol., 2019). Identifikovaných bolo osem oblastí orientovaných hlavne na priemysel, ako napríklad optimalizácia výroby, logistiky, riadenie kvality, prediktívna údržba, marketing a podpora, a ďalšie. Manuál obsahuje aj námety na využitie UI v strategických hospodárskych oblastiach, ako sú kybernetická bezpečnosť, zdravotníctvo, poľnohospodárstvo, verejná správa a životné prostredie.

Na národnej úrovni vzniklo na Slovensku Centrum pre umelú inteligenciu, ktoré organizuje rôzne aktivity v tejto oblasti. Od roku 2020 zastrešuje nezávislú národnú platformu AIslovakIA, ktorej ambíciou je rozvíjať excelenciu a spájať expertov a záujemcov o UI na Slovensku. Aktuálne najdôležitejšou oblasťou, kam sa orientujú aktivity Centra je zdravotníctvo. Súčasne Centrum realizuje marketingovo–komunikačnú kampaň na podporu nasadenia UI vo firmách.

3.3 Umelá inteligencia a vzdelávanie

Umelá inteligencia preniká postupne aj do vzdelávania a má veľký potenciál zmeniť metódy štúdia, výuky i prípravy naň. Doteraz vplyv UI nebol významnejší, no konkrétne s príchodom četbotov koncom roka 2022 sa situácia výrazne zmenila. Inteligentné četboty predstavujú začiatok novej epochy UI a jej dopadu na spoločnosť. Četboty ponúkajú de facto univerzálne použitie v rámci jazykovej domény, aj vo vzdelávaní. Mnohí študenti si už vypomáhajú četbotom ChatGPT pri tvorbe textov napr. záverečných prác, esejí alebo pri generovaní kusov zdrojového kódu pre počítače. Toto však vyvoláva potrebu prispôsobiť tomu metódy skúšania, pretože četbot sa tu stáva spoluautorom alebo jediným autorom. Zodpovednosť je na študentovi, ktorého prácu má učiteľ hodnotiť. Učitelia takisto môžu začať využívať tento nástroj, napríklad na zefektívnenie prípravy prednášok, alebo na prípravu testov (čo je časovo veľmi náročné).

Črtajú sa aj ďalšie možnosti ako využiť UI vo vzdelávaní. Spomeňme tri:

UI dokáže prispôsobiť vzdelávací obsah a skúsenosti individuálnym potrebám a štýlom učenia sa konkrétneho študenta, čomu sa hovorí *personalizované učenie*. Na základe analýzy údajov o výkone študentov môžu systémy UI (vytvorením si modelu užívateľa) odporučiť vhodné učebné materiály, cvičenia a tempo, ktoré študentovi umožnia učiť sa vlastnou rýchlosťou a spôsobmi, ktoré mu najviac vyhovujú.

UI môže podporovať *virtuálne laboratóriá a simulácie*, ktoré umožňujú študentom experimentovať a učiť sa v kontrolovanom di-

gitálnom prostredí, čo je obzvlášť užitočné pre predmety v oblastiach prírodných a technických vied. Virtuálna realita je sama osebe dôležitá technológia, v rámci ktorej vytvárané virtuálne svety sú použiteľné na rôzne konštruktívne účely, poskytujúc viacero výhod v porovnaní s fyzickým svetom.

UI dokáže *automatizovať hodnotenie úloh*, čím ušetrí učiteľom značný čas (najmä pri veľkých predmetoch) a študentom poskytne rýchlejšiu spätnú väzbu. To umožní pedagógom sústrediť sa viac na iné aktivity, ktoré si vyžadujú ľudskú odbornosť.

4 BUDÚCNOSŤ S UMELOU INTELIGENCIOU

Rozvoj umelej inteligencie sa nedá zastaviť, pravdepodobne ani spomaliť. Jedno je isté, na tieto dramatické očakávané zmeny sa spoločnosť musí pripraviť, či už legistatívne, psychologicky, pedagogicky alebo inak. Konkurencia vo vývoji UI bude určite narastať. Svedčí o tom celosvetový nárast investícií v tejto oblasti, ako aj fakt, že najbohatší muž planéty, Elon Musk, obchodný magnát a investor, majiteľ viacerých významných firiem, tento rok založil ďalšiu spoločnosť, xAI, zameranú na vývoj UI.

Expert na umelú inteligenciu, kolega Martin Takáč (2022) v zaujímavom rozhovore hovorí o budúcnosti UI a vidí to ideálne tak, že umelá inteligencia by bola všade okolo nás, ale takmer neviditeľná. Myslím si, že toto nebude úplne možné a nie je to ani žiaduce. V prípade vtelených systémov s UI bude zrejmé, s čím má človek do činenia. Napríklad v prípade inteligentných robotov, s ktorými budú ľudia častejšie interagovať. Ak budú autonómne autá dostatočne spoľahlivé, budú zrejme rozšírenou formou zdieľanej taxislužby riadenej umelou inteligenciou. Iné je to v prípade používania bežných nástrojov ako je osobný počítač, kde systém s UI beží na pozadí nejakej aplikácie. Napríklad, na sociálnych sieťach alebo pri googlení si bežný užívateľ neuvedomuje, že v pozadí je odporúčací systém alebo personalizovaný systém, ktorý užívateľovi podsúva to, čo očakáva alebo ho zaujíma. Toto ako vieme prispieva k vytváraniu sveta sociálnych bublín, ktoré

majú skôr negatívne dôsledky, lebo prispievajú k polarizácii spoločnosti. Toto je jeden z vážnych problémov UI, ktorému čelíme. Bolo by ideálne, keby v tomto smere UI robila skôr také aktivity, ktoré by viedli k spájaniu ľudí. Prakticky to je zrejme problém, lebo tu chýba atraktívny obchodný model. To by museli technologickí giganti chcieť inak uvažovať, v prospech civilizácie. Každá nová technológia má svoje riziká, ak sa dostane do nesprávnych rúk.

Neexistuje jednotný názor na to, ako bude UI vyzerať o 10 či 20 rokov (vyššie sme spomenuli singularitu). Je to pochopiteľné, pretože žijeme v exponenciálne sa zrýchľujúcej dobe technológií. Myslím si, že v snahe o odstránenie nedostatkov súčasných hlbokých modelov dôjde k paradigmatickému posunu smerom ku kognitívnej UI, ktorá má potenciál umožniť neurónovým sieťam rozumieť tomu, čo sa učia (Zhu a spol., 2020). Četboty už k tomuto smerujú – textové znalosti sa pridaním iných modalít (obraz, motorika) prepájajú s reálnym svetom, kde sa významy slov a fráz ukotvujú (Huang a spol., 2023). Aj dieťa vníma svet okolo seba, interaguje s ním pomocou svojho tela, učí sa slová jazyka v multimodálnom kontexte.

Je veľmi pravdepodobné, že využívanie UI sa bude rozširovať po celom svete a raz sa dotkne všetkých civilizácií. Postoje krajín k umelej inteligencii sú však rôzne. Vyspelé demokratické štáty už umelú inteligenciu začínajú regulovať, v autokratických krajinách však platia iné pravidlá účelovej regulácie, a známe sú príklady zneužívania UI napríklad na monitorovanie pohybu ľudí na základe rozpoznania ich tváre. Veľkým hráčom na svetovej scéne je Čína, ktorej systém sociálnych kreditov vyvoláva vo svete otázniky, pretože naráža na otázky etiky a slobody. Vojenské zneužitie umelej inteligencie je ešte väčšou hrozbou. Riziká teda existujú, rovnako ako aj príležitosti. Na pozitívne využitie umelej inteligencie na zlepšenie kvality života ľudí sa bohužiaľ nemôžeme automaticky spoliehať. Bude treba o to bojovať, rovnako ako o demokraciu.

LITERATÚRA

AI Act (2023). EU AI Act: first regulation on artificial intelligence. https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

Barredo-Arrieta A. a kol. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115.

Bečková I., Pócoš Š., Farkaš I. (2020). Computational analysis of robustness in neural network classifiers. V zborníku *Artificial Neural Networks and Machine Learning* (*ICANN*), Springer Nature Switzerland AG, 65-76.

Bengio Y., Courville A., Vincent P. (2013). Representation Learning: A Review and New Perspectives. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 35(8), 1798-1828.

Bieliková M. a kol. (2019). Analýza a návrh možností výskumu, vývoja a aplikácie umelej inteligencie na Slovensku. Dielo č. 2 – Manuál pre firmy na zavedenie umelej inteligencie. Vydavateľstvo STU v Bratislave. https://mirri.gov.sk/wp-content/uploads/2020/03/Dielo2-Manual.pdf

Biever, C. (2023). The Easy Intelligence Tests that AI Chatbots Fail. *Nature*, 619, 686-689.

Farkaš I. (2023). Dôveryhodnosť výpočtových modelov v umelej inteligencii a robotike. V zborníku *Kognícia a umelý život XXI*, Smolenice. Vydavateľstvo UK v Bratislave, 28-29.

Farkaš I., Cimrová B., Pócoš Š., Bečková I. (2022). Pozornosť ako biologicky inšpirovaný koncept pre vysvetliteľné, robustné a efektívne strojové učenie. V zborníku *Kognice a umělý život XX*, Třešť. Vydavatelství ČVUT v Praze, 34-38.

Fortunati, L., Esposito, A., Sarrica, M., and Ferrin, G. (2015). Children's knowledge and imaginary about robots. *International Journal of Social Robotics*, 7, 685-695.

Huang, S. a spol. (2023). Language is not all you need: Aligning perception with language models. *arXiv:2302.14045*.

Kvasnička V., Beňušková Ľ., Pospichal J., Farkaš I., Tiňo P., Kráľ A. (1997). *Úvod do teórie neurónových sietí*. IRIS, Bratislava.

Kurzweil R. (2005). *The Singularity is Near*. New York: Viking Books.

McCarthy J. (1999). What is AI? Basic questions. http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html

Mnih V. a spol. (2015). Human-level control through deep reinforcement learning. *Nature*, 518:529-542.

Montavon G., Samek W. , Müller K.-R. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73, 1-15.

OpenAI (2023). GPT-4 technical report. *https://arxiv.org/abs/2303.08774*.

Pinker S. (2022). *Racionalita: Čo je to. Kam sa podela. Na čo nám je*. Tatran.

Pócoš Š., Bečková I., Farkaš I. (2022). Examining the proximity of adversarial examples to class manifolds in deep networks. V zborníku *International Conference on Artificial Neural Networks (ICANN)*, Springer, 645-656.

Popper M. (1986). CODEX: an expert system for medical applications. *Časopis lékařů českých*. 125(2), 40-4.

Russell S.,Norvig P. (2010). *Artificial Intelligence: A Modern Approach* (3. vyd.). Prentice Hall.

Schmidhuber J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85-117.

Sinčák P., Andrejková G. (1986). *Neurónové siete: Inžiniersky prístup* (1. a 2. diel). Vydavateľstvo Elfa, Košice.

Šefránek J., Takáč M., Farkaš I. (2008). Vznik inteligencie v umelých systémoch. V knihe Magdolen D. (zost.), *Hmota, život, inteligencia: Vznik*. Vydavateľstvo VEDA, Bratislava. 245-270.

Takáč M. (2022). Umelá inteligencia prinesie v tomto storočí prudké zmeny. Zíde sa nám flexibilita a odolnosť voči stresu, vraví expert. *Denník N*, 30.12.2022.

Takáč M. (2023). Čo chýba ChatGPT k tomu, aby rozumel, čo robí? V zborníku *Kognícia a umelý život XXI*. Vydavateľstvo UK v Bratislave, 74-75.

Tomasello M. (2001). *The Cultural Origins of Human Cognition*. Harvard University Press.

Zhao W.X. a spol. (2023). A Survey of Large Language Models. *arXiv: 2303.18223*

Zhu Y. a spol. (2020). Dark, Beyond Deep: A Paradigm Shift to Cognitive AI with Humanlike Common Sense. *Engineering*, 6(3), 310-345.