

NEZÁVISLÁ

EXPERTNÁ SKUPINA NA VYSOKEJ ÚROVNI PRE UMELÚ INTELIGENCIU

ZRIADENÁ EURÓPSKOU KOMISIOU V JÚNI 2018



VYMEDZENIE POJMU UMELEJ INTELIGENCIE: HLAVNÉ SCHOPNOSTI A DISCIPLÍNY

Vymedzenie vypracované na účely výstupov skupiny

Vymedzenie pojmu umelej inteligencie: hlavné schopnosti a vedecké disciplíny

Expertná skupina na vysokej úrovni pre umelú inteligenciu

Vyhlásenie o odmietnutí zodpovednosti a použitie tohto dokumentu: Tento opis a vymedzenie schopnosti umelej inteligencie a oblastí výskumu predstavuje veľmi hrubé zjednodušenie aktuálneho stavu. Zámerom tohto dokumentu nie je presne a komplexne vymedziť všetky techniky a schopnosti umelej inteligencie, ale súhrnne opísať spoločné chápanie tohto odboru, ktoré expertná skupina na vysokej úrovni používa vo svojich výstupoch. Dúfame však, že tento dokument poslúži aj ako užitočné študijné východisko pre ľudí, ktorí nie sú odborníkmi na umelú inteligenciu a ktorí naň neskôr môžu nadviazať rozsiahlejším a hlbším zamyslením sa nad umelou inteligenciou s cieľom získať podrobnejšie znalosti o tejto disciplíne a technológii.

Expertná skupina na vysokej úrovni pre umelú inteligenciu je nezávislá expertná skupina, ktorú v júni 2018 zriadila Európska komisia.

Kontakt Nathalie Smuha – koordinátorka AI HLEG
E-mail CNECT-HLG-AI@ec.europa.eu

Európska komisia
B-1049 Brussels

Dokument zverejnený X. apríla 2019.

Prvý návrh tohto dokumentu bol vydaný 18. decembra 2018 spoločne s prvým návrhom etických usmernení pre dôveryhodnú umelú inteligenciu expertnej skupiny na vysokej úrovni pre umelú inteligenciu. Dokument bol revidovaný so zreteľom na pripomienky získané od Európskej aliancie pre umelú inteligenciu a z otvorených konzultácií o návrhu usmernení. Chceli by sme výslovne a srdečne poďakovať všetkým, ktorí poskytli svoju spätnú väzbu k prvému návrhu dokumentu.

Európska komisia ani žiadna osoba konajúca v mene Komisie nie je zodpovedná za možné použitie uvedených informácií. Obsah tohto pracovného dokumentu je výlučnou zodpovednosťou expertnej skupiny na vysokej úrovni pre umelú inteligenciu (AI HLEG). Hoci sa na vypracovaní tohto dokumentu podieľali zamestnanci Komisie, názory vyjadrené v tomto dokumente odzrkadľujú stanovisko AI HLEG a za žiadnych okolností ich nemožno považovať za vyjadrenie oficiálneho stanoviska Európskej komisie.

Viac informácií o expertnej skupine na vysokej úrovni pre umelú inteligenciu je k dispozícii online (<https://ec.europa.eu/digital-single-maret/en/high-level-expert-group-artificial-intelligence>).

Pravidlá opakovaného použitia dokumentov Európskej komisie sú upravené rozhodnutím 2011/833/EÚ (Ú. v. EÚ L 330, 14.12.2011, s. 39). Na akékoľvek použitie alebo reprodukciu fotografií alebo iného materiálu, na ktoré sa nevzťahujú autorské práva EÚ, je potrebné povolenie priamo od držiteľov príslušných autorských práv.

VYMEDZENIE POJMU UMELEJ INTELEGENCIE: HLAVNÉ SCHOPNOSTI A VEDECKÉ DISCIPLÍNY

Vychádzame z tohto vymedzenia pojmu umelej inteligencie navrhnutého v oznámení Európskej komisie o umelej inteligencii¹:

„Umelá inteligencia sú systémy, ktoré vykazujú inteligentné správanie tým, že analyzujú okolité prostredie a podnikajú kroky – s istou mierou samostatnosti – na dosiahnutie konkrétnych cieľov.

Systémy umelej inteligencie môžu byť založené výlučne na softvéri a pôsobiť vo virtuálnom svete (napr. hlasoví asistenti, softvér na analýzu fotografií, vyhľadávače, systémy rozpoznávania hlasu a tváre), ale umelá inteligencia môže byť aj súčasťou hardvérových zariadení (napr. vyspelé roboty, autonómne vozidlá, bezpilotné vzdušné prostriedky alebo aplikácie internetu vecí).“

V tomto dokumente sme toto vymedzenie rozšírili, aby sme objasnili určité aspekty umelej inteligencie ako vedeckej disciplíny a technológie s cieľom predísť nedorozumeniam, dosiahnuť jednotné spoločné porozumenie umelej inteligencie, ktorú môžu úspešne používať aj osoby, ktoré nie sú odborníkmi na umelú inteligenciu, a s cieľom poskytnúť užitočné informácie, ktoré sa môžu použiť v diskusii o etických usmerneniach pre umelú inteligenciu a o odporúčaníach pre politiky v oblasti umelej inteligencie.

1. Systémy umelej inteligencie

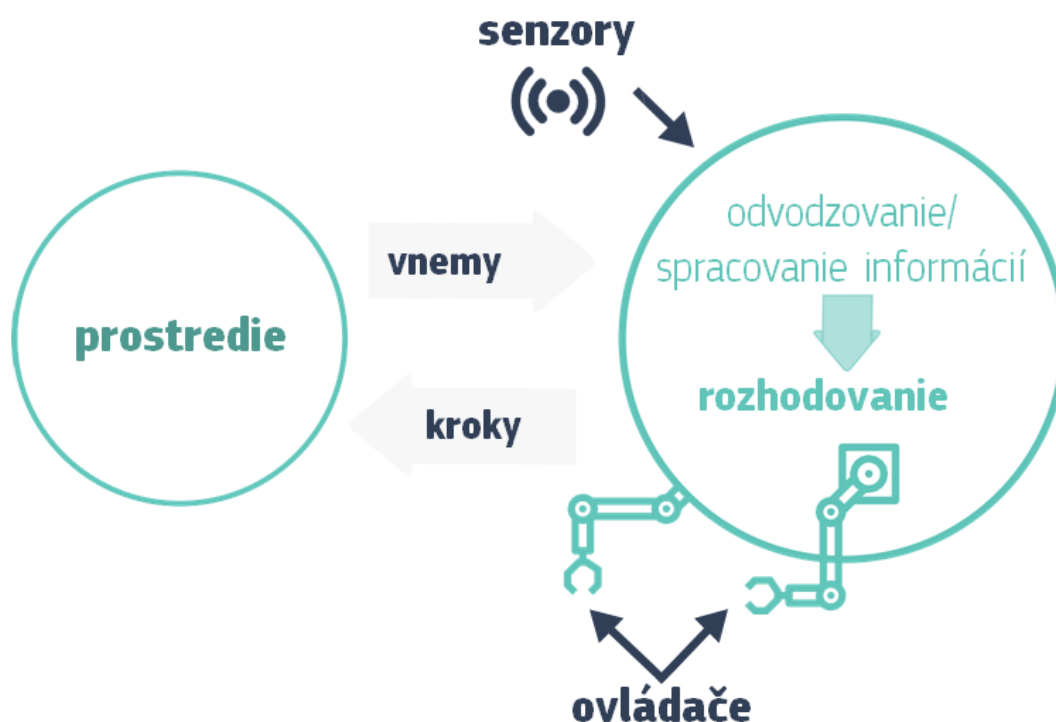
V termíne umelá inteligencia sa výslovne spomína pojem inteligencia. Keďže však je inteligencia (v prípade strojov aj v prípade ľudí) nejasný pojem, hoci je predmetom podrobného štúdia psychológov, biológov a neurovedcov, výskumní pracovníci v oblasti umelej inteligencie väčšinou používajú pojem „racionálnosť“. Tento pojem označuje schopnosť vybrať najvhodnejší krok na dosiahnutie určitého cieľa s ohľadom na isté kritériá, ktoré sa majú optimalizovať, a na dostupné zdroje. Racionálnosť, samozrejme, nie je jediným prvkom koncepcie inteligencie, tvorí však jej podstatnú časť.

V tomto dokumente budeme používať termín *systém umelej inteligencie* vo význame akéhokoľvek komponentu, softvéru a/alebo hardvéru založeného na umelej inteligencii. Systémy umelej inteligencie naozaj zvyčajne bývajú *zabudované* do väčších systémov ako ich zložky, menej často ide o samostatné systémy.

Systém umelej inteligencie je tak predovšetkým racionálny, ako sa uvádza v jednej z najpoužívanějších učebníc o umelej inteligencii². Ako však systém umelej inteligencie dosiahne racionálnosť? Ako sa uvádza v prvej vete citovaného pracovného vymedzenia pojmu umelá inteligencia, racionálnosť dosahuje vnímaním okolitého prostredia, ktorého súčasťou je prostredníctvom senzorov, a takto zhromažďuje a interpretuje údaje, odvodzuje z vnímaných údajov alebo spracúva informácie odvodené z týchto údajov a rozhoduje o najlepšom kroku a potom na základe toho koná prostredníctvom určitých ovládačov, čím môže upravovať svoje prostredie. Systémy umelej inteligencie môžu buď používať symbolické pravidlá, alebo sa naučiť numerický model, a takisto môžu upraviť svoje správanie na základe analýzy vplyvu, aký malo ich predchádzajúce konanie na ich prostredie. Pomôckou by mohla byť ilustrácia systému umelej inteligencie na obrázku 1.

¹ Oznámenie Komisie Európskemu parlamentu, Európskej rade, Rade, Európskemu hospodárskemu a sociálnemu výboru a Výboru regiónov o umelej inteligencii pre Európu, Brusel, 25. 4. 2018, COM(2018) 237 final.

² *Artificial Intelligence: A Modern Approach* (Umelá inteligencia: moderný prístup), Russell, S., a Norvig, P., Prentice Hall, 3. vydanie, 2009.



Obrázok 1: Schematické znázornenie systému umelej inteligencie.

Senzory a percepcia. Na obrázku 1 sú senzory systému znázornené symbolom pre wifi. V praxi môže ísť o kamery, mikrofóny, klávesnicu, webové sídlo alebo iné vstupné zariadenia, ako aj o senzory na meranie fyzikálnych veličín (napr. teploty, tlaku, vzdialenosti, sily/krútiaceho momentu, dotykové snímače). Vo všeobecnosti systémy umelej inteligencie musíme vybaviť senzormi, ktoré sú vhodné na sledovanie údajov nachádzajúcich sa v okolitom prostredí a relevantných pre cieľ, ktorý systému umelej inteligencie určil jeho ľudský tvorca. Ak napríklad chceme vytvoriť systém umelej inteligencie, ktorý bude automaticky čistiť podlahu miestnosti, keď sa zašpiní, senzormi by mohli byť kamery na snímánie podlahy.

Pokiaľ ide o zhromaždené údaje, často je užitočné rozlišovať medzi štruktúrovanými a neštruktúrovanými údajmi. *Štruktúrované údaje* sú údaje, ktoré sú usporiadané podľa vopred stanovených modelov (napríklad v relačnej databáze), kým *neštruktúrované údaje* nie sú usporiadané známym spôsobom (napríklad v obrázku alebo v textovom úryvku).

Odvodzovanie/spracovanie informácií a rozhodovanie. Jadro systému umelej inteligencie tvorí jeho modul na odvodzovanie/spracovanie informácií, ktorý na vstupe prijíma údaje prichádzajúce zo sensorov a navrhuje, aký krok sa má uskutočniť s ohľadom na cieľ, ktorý sa má dosiahnuť. To znamená, že údaje zhromaždené senzormi sa musia premeniť na informáciu, ktorej tento modul rozumie. Ak budeme pokračovať v našom príklade upratovacieho systému umelej inteligencie, kamera zabezpečí snímku podlahy pre modul na odvodzovanie/spracovanie informácií a tento modul musí rozhodnúť, či má vyčistiť podlahu alebo nie (t. j. aký je najvhodnejší krok na dosiahnutie požadovaného cieľa). Hoci sa nám, ľuďom, môže zdať jednoduché na základe obrázku podlahy dospieť k rozhodnutiu o tom, či potrebuje vyčistiť, pre stroj toto rozhodnutie také jednoduché nie je, pretože obrázok je len reťazec núl a jednotiek. Modul na odvodzovanie/spracovanie informácií preto musí:

1. Interpretovať obrázok s cieľom určiť, či je podlaha čistá. To vo všeobecnosti znamená, že musí byť schopný premeniť údaje na informácie a modelovať tieto informácie výstižným spôsobom, ktorý by však mal zahŕňať všetky podstatné časti údajov (v tomto prípade to, či je podlaha čistá).
2. Odvodzovať z týchto poznatkov alebo spracovať uvedené informácie s cieľom vytvoriť numerický model (čiže matematický vzorec), aby mohol rozhodnúť o najvhodnejšom kroku. V tomto príklade, ak sa z obrázku odvodí

informácia, že podlaha je špinavá, najvhodnejším krokom bude aktivácia čistenia, v opačnom prípade ním je nečinnosť.

Pripomíname, že pojem „rozhodnutie“ treba chápať vo všeobecnosti ako akýkoľvek úkon výberu kroku a nevyhnutne neznamená, že systémy umelej inteligencie sú úplne samostatné. Rozhodnutím môže byť aj výber odporúčania, ktoré sa má poskytnúť človeku, ktorý prijme konečné rozhodnutie.

Uvedenie do činnosti. Vo chvíli, keď sa prijme rozhodnutie o kroku, systém umelej inteligencie je pripravený ho vykonať prostredníctvom ovládačov, ktorými disponuje. Na obrázku 1 sú ovládače znázornené ako kľbové ramená, ovládače však nemusia byť fyzické. Môžu byť aj softvérové. V našom príklade upratovacieho systému by systém umelej inteligencie mohol vyslať signál, ktorým sa aktivuje vysávač, ak je krokom čistenie podlahy. V inom príklade konverzačný systém (čiže chatbot) koná tak, že vytvára texty, ktorými odpovedá na vyjadrenia používateľa.

Vykonaný krok pravdepodobne zmení prostredie, takže nabudúce, keď systém bude musieť znovu použiť svoje senzory, zo zmeneného prostredia bude možno vnímať odlišné informácie.

Racionálne systémy umelej inteligencie nie vždy v záujme svojho cieľa zvolia najvhodnejší krok, takže dosiahnu iba *obmedzenú racionalitu*, a to z dôvodu obmedzení zdrojov, ako sú čas alebo výpočtová kapacita.

Racionálne systémy umelej inteligencie sú úplne najzákladnejšou verziou systémov umelej inteligencie. Menia prostredie, v priebehu času však neprispôbujú svoje správanie tak, aby dosiahli svoj cieľ lepším spôsobom. *Učiaci sa racionálny systém* je racionálny systém, ktorý po prijatí kroku vyhodnocuje nový stav prostredia (pomocou percepcie), aby mohol určiť mieru úspešnosti svojho kroku, a následne prispôbi svoje pravidlá odvodzovania a metódy rozhodovania.

2. Umelá inteligencia ako vedecká disciplína

Uvedený opis je veľmi jednoduchý abstraktný opis systému umelej inteligencie prostredníctvom jeho troch hlavných schopností – percepcie, odvodzovania/rozhodovania a uvedenia do činnosti. Stačí však na to, aby sme mohli predstaviť väčšinu techník a podoblastí umelej inteligencie, ktoré sa v súčasnosti používajú na budovanie systémov umelej inteligencie, a aby sme im mohli porozumieť, keďže všetky tieto techniky a podoblasti súvisia s rozličnými schopnosťami systémov. Vo všeobecnosti možno všetky tieto techniky rozdeliť na dve hlavné skupiny, ktoré súvisia so schopnosťou *odvodzovania* a *učenia sa*. Robotika je ďalšou veľmi dôležitou disciplínou.

Odvodzovanie a rozhodovanie. Do tejto skupiny techník patrí reprezentácia a odvodzovanie poznatkov, plánovanie, programovanie, vyhľadávanie a optimalizácia. Tieto techniky umožňujú odvodzovanie z údajov pochádzajúcich zo senzorov. Na tento účel je potrebné premeniť údaje na poznatky, takže jedna oblasť umelej inteligencie sa zaoberá tým, ako čo najlepšie modelovať tieto poznatky (*reprezentácia poznatkov*). Ďalším krokom po dokončení modelovania poznatkov je odvodiť z nich logický záver (*odvodzovanie z poznatkov*), čo zahŕňa vyvodzovanie záverov prostredníctvom symbolických pravidiel, činnosti *plánovania* a *programovania*, *vyhľadávanie* prostredníctvom rozsiahlej množiny riešení a *optimalizácia* spomedzi všetkých možných riešení problému. Poslednou fázou je rozhodnúť, aký krok sa má vykonať. Časť systému umelej inteligencie týkajúca sa odvodzovania/rozhodovania zvyčajne býva veľmi zložitá a vyžaduje si kombináciu niekoľkých uvedených techník.

Učenie sa. Táto skupina techník zahŕňa strojové učenie, neurónové siete, hĺbkové učenie, rozhodovacie stromy a veľa ďalších techník učenia sa. Tieto techniky systému umelej inteligencie umožňujú učiť sa, ako riešiť problémy, ktoré nie je možné presne stanoviť alebo ktorých metódu riešenia nemožno opísať pomocou symbolických pravidiel odvodzovania. Medzi príklady takýchto problémov patria tie, ktoré sa zaoberajú perцепčnými schopnosťami, ako je porozumenie *reči* a *jazyka*, ako aj *počítačové videnie* alebo *predikcia správania*. Treba upozorniť, že tieto problémy sa zdajú jednoduché, pretože sú zvyčajne vlastne jednoduché pre ľudí. Pre systémy umelej inteligencie však nie sú až také jednoduché, keďže tie sa nemôžu spoliehať na odvodzovanie na základe zdravého rozumu (aspoň zatiaľ nie) a sú obzvlášť náročné, ak systém musí interpretovať neštruktúrované údaje. V týchto prípadoch prídu vhod techniky, ktoré sa riadia prístupom *strojového učenia*. Techniky strojového učenia sa však môžu použiť na oveľa viac úloh než

iba na percepciu. Vytvárajú numerický model (čiže matematický vzorec), ktorý sa používa na výpočet rozhodnutia na základe údajov.

Strojové učenie má niekoľko podôb. Najrozšírenejšie prístupy sú *učenie s učiteľom*, *učenie bez učiteľa* a *učenie posilňovaním*.

Pri strojom učení s učiteľom sa systému namiesto toho, aby sa mu zadali pravidlá správania, poskytnú príklady správania vstup-výstup s očakávaním, že systém bude schopný zovšeobecniť na základe príkladov (zvyčajne opisom minulosti) a správať sa správne aj v situáciách, ktoré neboli uvedené medzi príkladmi (ku ktorým by mohlo dôjsť v budúcnosti). V našom príklade by sme systému poskytli množstvo príkladov obrázkov s podlahou a ich príslušnou interpretáciou (t. j. či je podlaha na danom obrázku čistá). Ak do systému zadáme dostatočný počet príkladov, ktoré sú dostatočne rozmanité a reprezentatívne pre väčšinu situácií, systém bude schopný prostredníctvom algoritmu strojového učenia zovšeobecniť tak, že bude vedieť správne interpretovať aj obrázky podláh, s ktorými sa ešte nestretol. Niektoré prístupy k strojovému učeniu využívajú algoritmy založené na koncepcii *neurónových sietí*, ktoré sú voľne inšpirované ľudským mozgom v tom, že systém tvorí sieť malých procesorov (podobne ako naše neuróny) s množstvom prepojení ohodnotených váhovým koeficientom medzi nimi. Vstupom pre neurónovú sieť sú údaje pochádzajúce zo senzorov (v našom príklade ide o obrázok podlahy) a výstupom je interpretácia obrázku (v našom príklade to, či je podlaha čistá). Počas analýzy príkladov (fáza *výcviku* siete) sa váhové koeficienty prepojení prispôbobia tak, aby čo najviac zodpovedali tvrdeniam dostupných príkladov (čiže s cieľom minimalizovať chyby medzi očakávaným výstupom a výstupom, ktorý je výsledkom výpočtov siete). Na konci fázy výcviku nasleduje testovacia fáza správania neurónovej siete, počas ktorej sa na základe príkladov, ktoré systém nikdy nevidel, overí, či sa systém úlohu naučil správne.

Treba si všimnúť, že tento prístup (ako všetky techniky strojového učenia) má vždy určitý percentuálny podiel chybovosti, i keď je tento podiel nízky. Zásadným pojmom je teda *presnosť*, čo je miera toho, aké vysoké je percento správnych odpovedí.

Existuje niekoľko druhov neurónových sietí a prístupov strojového učenia, spomedzi ktorých je v súčasnosti jedným z najúspešnejších *hlbkové učenie*. Tento prístup poukazuje na to, že neurónová sieť má viacero vrstiev medzi vstupom a výstupom, ktoré umožňujú naučiť sa celkový vzťah medzi vstupom a výstupom v sérii nadväzujúcich krokov. V dôsledku toho je celkový prístup presnejší a vyžaduje si menej usmerňovania zo strany ľudí.

Neurónové siete sú len jedným z nástrojov strojového učenia, existuje však mnoho ďalších nástrojov s odlišnými vlastnosťami: náhodné lesy a boostingové stromy, metódy tvorby klastrov, faktorizácia matice atď.

Ďalší užitočný druh prístupu strojového učenia sa nazýva *učenie posilňovaním*. Pri tomto prístupe sa systému umelej inteligencie ponecháva sloboda prijímať vlastné rozhodnutia a pri každom rozhodnutí sa mu poskytne signál odmeny, ktorý ho informuje o tom, či spravil dobré alebo zlé rozhodnutie. Cieľom systému je časom maximalizovať príjem pozitívnych odmien. Tento prístup sa používa napríklad v rámci systému na vytváranie odporúčaní (ako je niekoľko online systémov na vytváranie odporúčaní, ktoré používateľom radia, čo by mohli chcieť kúpiť) alebo v marketingu.

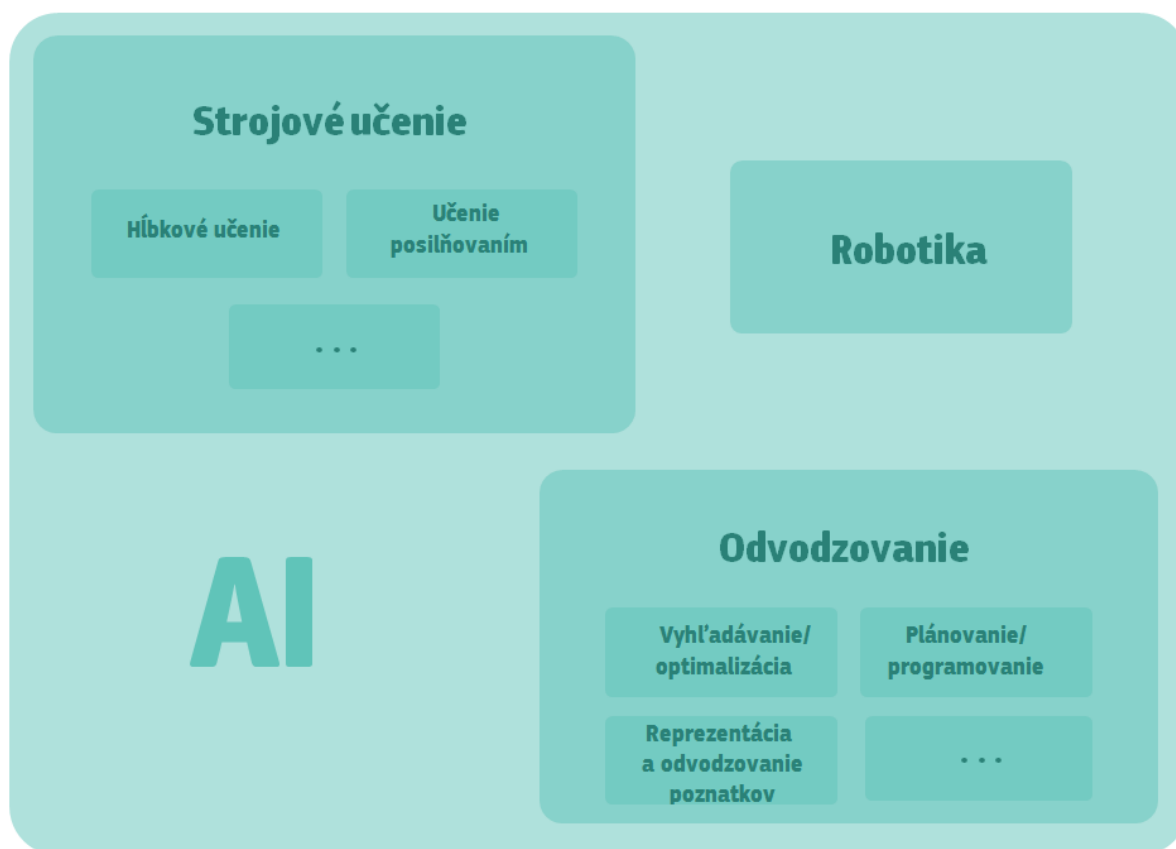
Prístupy strojového učenia nie sú užitočné iba v úlohách týkajúcich sa percepcie, ako je videnie alebo porozumenie textu, ale aj vo všetkých úlohách, ktoré je ťažko vymedziť a ktoré nemožno úplne opísať pomocou symbolických pravidiel správania.

Treba upozorniť na rozdiel medzi prístupmi strojového učenia k učeniu novej úlohy, ktorú nie je možné vhodne symbolicky opísať, a učiacimi sa racionálnymi subjektmi (uvedené v predchádzajúcom oddiele), ktoré časom prispôbujú svoje správanie tak, aby mohli lepšie dosiahnuť daný cieľ. Tieto dve techniky sa môžu prekrývať alebo medzi nimi môže byť spolupráca, nie sú však nevyhnutne rovnaké.

Robotika. Robotiku možno definovať ako „umelú inteligenciu v akcii vo fyzickom svete“ (takisto sa označuje ako *stelesnená umelá inteligencia*). Robot je fyzický stroj, ktorý sa musí vyrovnávať s dynamikou, neistotami a zložitou fyzikou sveta. Schopnosti percepcie, odvodzovania, konania, učenia sa, ako aj interakcie s inými systémami sú zvyčajne začlenené do architektúry riadenia robotického systému. V konštrukcii a prevádzke robotov okrem umelej inteligencie zohrávajú úlohu aj iné disciplíny, ako je strojárstvo a teória riadenia. Medzi príklady robotov patria

robotické manipulátory, autonómne vozidlá (napr. automobily, bezpilotné vzdušné prostriedky, lietajúce taxíky), humanoidné roboty, robotické vysávače atď.

Na obrázku 2 je znázornená väčšina uvedených podoblastí umelej inteligencie, ako aj ich vzájomný vzťah. Dôležité je však upozorniť na to, že umelá inteligencia je oveľa zložitejšia, než vyplýva z tohto obrázka, keďže zahŕňa množstvo ďalších podoblastí a techník. Okrem toho, ako už bolo uvedené, robotika sa opiera aj o techniky, ktoré nepatria do oblasti umelej inteligencie. Domnievame sa však, že to stačí na úspešné poskytovanie informácií pre vzájomnú výmenu informácií, zvyšovanie informovanosti a diskusie o umelej inteligencii, etike umelej inteligencie a politikách v oblasti umelej inteligencie, ktoré sa musia uskutočniť v samotnej multidisciplinárnej a mnohostrannej expertnej skupine na vysokej úrovni.



Obrázok 2: Zjednodušený prehľad o podoblastiach umelej inteligencie a ich vzájomných vzťahoch. Strojové učenie aj odvodzovanie obsahuje mnoho ďalších techník, pričom robotika zahŕňa techniky, ktoré nie sú súčasťou umelej inteligencie. Celá umelá inteligencia patrí do disciplíny počítačovej vedy.

3. Ďalšie dôležité pojmy a otázky súvisiace s umelou inteligenciou

Obmedzená (alebo slabá) a všeobecná (alebo silná) umelá inteligencia. Všeobecný systém umelej inteligencie má byť systémom, ktorý môže vykonávať väčšinu činností, ktoré môžu vykonávať ľudia. Obmedzené systémy umelej inteligencie sú naopak systémy, ktoré môžu vykonávať jednu alebo niekoľko osobitných úloh. V súčasnosti zavedené systémy umelej inteligencie sú príkladmi obmedzenej umelej inteligencie. V počiatkoch umelej inteligencie výskumní pracovníci používali odlišnú terminológiu (slabá a silná umelá inteligencia). Stále existuje mnoho otvorených etických, vedeckých a technologických problémov pre vytvorenie schopností, ktoré sú potrebné na dosiahnutie všeobecnej umelej inteligencie, ako je odvodzovanie na základe zdravého rozumu, sebauvedomovanie a schopnosť stroja určiť svoj vlastný účel.

Otázky týkajúce sa údajov a zaujatosti. Mnohé systémy umelej inteligencie, ako sú systémy obsahujúce zložky strojového učenia s učiteľom, sa spoliehajú na obrovské množstvá údajov na to, aby mohli fungovať dobre. Preto je dôležité chápať, ako údaje ovplyvňujú správanie systému umelej inteligencie. Napríklad ak sa údaje určené na výcvik vyznačujú zaujatosťou (čiže nie sú dostatočne vyvážené alebo inkluzívne), systém umelej inteligencie vycvičený na základe týchto údajov nebude môcť správne zovšeobecňovať a pravdepodobne bude prijímať nespravodlivé rozhodnutia, ktoré môžu uprednostňovať niektoré skupiny pred inými. Komunita okolo umelej inteligencie v poslednom čase pracuje na metódach, ktoré by odhaľovali a zmierňovali zaujatosť v súboroch údajov určených na výcvik, ako aj v ďalších častiach systému umelej inteligencie.

Umelá inteligencia ako čierna skrinka a vysvetliteľnosť. Hoci sú niektoré techniky strojového učenia veľmi úspešné z hľadiska presnosti, z hľadiska chápania nie je veľmi jasné, ako prijímajú rozhodnutia. Pojem *umelá inteligencia ako čierna skrinka* označuje scenáre, pri ktorých nie je možné vysledovať dôvod istých rozhodnutí. Vysvetliteľnosť je vlastnosť tých systémov umelej inteligencie, ktoré namiesto toho môžu poskytnúť istú formu vysvetlenia svojich krokov.

Umelá inteligencia zameraná na cieľ. Súčasné systémy umelej inteligencie sú zamerané na cieľ, čo znamená, že opis cieľa, ktorý majú dosiahnuť, dostávajú od ľudí a používajú určité techniky na jeho dosiahnutie. Nestanovujú si vlastné ciele. Niektoré systémy umelej inteligencie (ako sú systémy založené na určitých technikách strojového učenia) však môžu mať viac slobody rozhodnúť sa, ktorý postup si na dosiahnutie daného cieľa vyberú.

4. Aktualizované vymedzenie pojmu umelej inteligencie

Navrhujeme používať toto aktualizované vymedzenie pojmu umelej inteligencie:

„Systémy umelej inteligencie sú softvérové (a prípadne aj hardvérové) systémy navrhnuté ľuďmi³, ktoré vzhľadom na komplexný cieľ konajú vo fyzickom alebo digitálnom rozmere tak, že vnímajú svoje prostredie prostredníctvom získavania údajov, interpretácie zhromaždených štruktúrovaných alebo neštruktúrovaných údajov, odvodzovania z poznatkov alebo spracúvania informácií odvodených z týchto údajov a že rozhodujú o najlepších krokoch, ktoré sa majú vykonať na dosiahnutie daného cieľa. Systémy umelej inteligencie môžu buď používať symbolické pravidlá, alebo sa naučiť numerický model, a takisto môžu upraviť svoje správanie na základe analýzy vplyvu, aký malo ich predchádzajúce konanie na ich prostredie.

Umelá inteligencia ako vedecká disciplína obsahuje niekoľko prístupov a techník, ako je strojové učenie (ktorého konkrétnymi príkladmi sú hĺbkové učenie a učenie posilňovaním), strojové odvodzovanie (ktorého súčasťou je plánovanie, programovanie, reprezentácia a odvodzovanie poznatkov, vyhľadávanie a optimalizácia) a robotika (do ktorej patrí kontrola, percepcia, senzory a ovládače, ako aj začlenenie všetkých ostatných techník do kyberneticko-fyzických systémov).“

a odvolávať sa na tento dokument ako na zdroj doplňujúcich informácií v záujme podpory tohto vymedzenia.

³ Ľudia navrhujú systémy umelej inteligencie priamo, môžu však použiť aj technológie umelej inteligencie na optimalizáciu ich dizajnu.

Tento dokument vypracovali členovia expertnej skupiny na vysokej úrovni pre umelú inteligenciu

uvedení ďalej v abecednom poradí

Pekka Ala-Pietilä, predseda expertnej skupiny na vysokej úrovni pre umelú inteligenciu AI Finland, Huhtamaki, Sanoma	Pierre Lucas Orgalim – Európske technologické odvetvia
Wilhelm Bauer Fraunhofer	Ieva Martinkenaite Telenor
Urs Bergmann Zalando	Thomas Metzinger JGU Mainz a Európske združenie univerzít
Mária Bieliková Slovenská technická univerzita v Bratislave	Catelijne Muller ALLAI Holandsko a EHSV
Cecilia Bonefeld-Dahl DigitalEurope	Markus Noga SAP
Yann Bonnet ANSSI	Barry O'Sullivan, podpredseda AI HLEG University College Cork
Loubna Bouarfa OKRA	Ursula Pacht BEUC
Stéphane Brunessaux Airbus	Nicolas Petit Univerzita v Liège (L'Université de Liège)
Raja Chatila Iniciatíva IEEE pre etiku inteligentných/autonómnych systémov a univerzita Sorbonna (Sorbonne Université)	Christoph Peylo Bosch
Mark Coeckelbergh Viedenská univerzita (Universität Wien)	Iris Plöger BDI
Virginia Dignum Univerzita v Umeå (Umeå universitet)	Stefano Quintarelli Garden Ventures
Luciano Floridi Univerzita v Oxforde (University of Oxford)	Andrea Renda College of Europe a CEPS
Jean-Francois Gagné Element AI	Francesca Rossi* IBM
Chiara Giovannini ANEC	Cristina San José Európska banková federácia
Joanna Goodey Agentúra pre základné práva	George Sharkov Digital SME Alliance
Sami Haddadin Mníchovská škola robotiky a strojovej inteligencie (Munich School of Robotics and MI)	Philipp Slusallek Nemecké výskumné centrum pre umelú inteligenciu (DFKI)
Gry Hasselbalch The thinkdotank DataEthics a Univerzita v Kodani (Københavns Universitet)	Françoise Soulié Fogelman Konzultantka pre oblasť umelej inteligencie
Fredrik Heintz Univerzita v Linköpingu (Linköpings universitet)	Saskia Steinacker Bayer
Fanny Hidvegi Access Now	Jaan Tallinn Ambient Sound Investment
Eric Hilgendorf Univerzita vo Würzburgu (Universität Würzburg)	Thierry Tingaud STMicroelectronics
Klaus Höckner Hilfsgemeinschaft der Blinden und Sehschwachen	Jakob Uszkoreit Google
Mari-Noëlle Jégo-Laveissière Orange	Aimee Van Wynsberghe TU Delft
Leo Kärkkäinen Nokia Bell Labs	Thiébaut Weber ETUC
Sabine Theresia Köszegi TU Wien	Cecile Wendling AXA
Robert Kroplewski Právny zástupca a poradca poľskej vlády	Karen Yeung Univerzita v Birminghame (University of Birmingham)
Elisabeth Ling RELX	

* Francesca Rossi vystupovala ako spravodajkyňa k tomuto dokumentu.

