

IRELEVANTNOSŤ TURINGOVHO TESTU V SÚČASNOM HLBOKOM UČENÍ

ONDREJ HRIADEL

Katedra filozofie a aplikovanej filozofie, Filozofická fakulta
Univerzity sv. Cyrila a Metoda v Trnave, Slovensko, ondro.hriadel@gmail.com

PŮVODNÍ VĚDECKÁ PRÁCE ▪ OBDRŽENO: 9/9/2021 ▪ PŘIJATO: 23/11/2021

Abstrakt: Úlohou umelej inteligencie (UI) v Turingovom teste je imitovať človeka do takej miery, aby vyšetrovateľ nebol schopný rozlíšiť stroj od človeka. S príchodom hlbokého učenia (DL) (podkategória UI) sa však situácia mení, pretože sa tieto systémy namiesto simulovania ľudskej inteligencie zameriavajú na riešenie konkrétnych problémov. Z dôvodu, že tieto umelé systémy nesimulujú ľudskú inteligenciu, sa otvára otázka, či nie je Turingov test v problematike hlbokého učenia irelevantný. Na problém sa je možné pozrieť v troch častiach. Po prvé, sa je potrebné zamerať na aplikačné využitie Turingovho testu v Loebnerovej cene, v ktorej sú kladené otázky zamerané na aspekty ľudskej inteligencie – učenie, usudzovanie a porozumenie. Po druhé, je možné považovať za problém, že sa v Turingovom teste rozumie pod inteligenciou iba všeobecná ľudská inteligencia. Keďže ani DL touto formou inteligencie nedisponuje, je možné bez pochyb označiť túto UI za neinteligentnú? Nakoniec je otázne, či by vlastne malo zmysel, aby účelovo zameraná UI, akou je DL, absolvovala Turingov test, nakoľko samotný test žiadne ďalšie poznatky o analýze problémov alebo inteligencii neprináša.

Kľúčové slová: hlboké učenie (deep learning, DL); umelá inteligencia (UI); Turingov test; inteligencia; učenie; natrénovanie dát; jazyk

THE IRRELEVANCE OF THE TURING TEST IN CURRENT DEEP LEARNING

Abstract: The role of artificial intelligence in the Turing test is to imitate human beings to such an extent that people will not realize it is a machine. With the rise of deep learning (a subcategory of AI), the situation is changing rapidly as the new systems do not focus on imitating human intelligence but emphasize thorough solutions to specific issues. The main difference between predefined AI and deep learning (DL) is that these systems are self-learning and have verifiable results. Firstly, we need to analyse the application of the Turing test in the Loebner Prize because, there, the primary emphasis is on aspects of human intelligence – learning, reasoning and understanding. Secondly, in the Turing test, only general intelligence is considered, and this can be questionable. If DL does not possess this form of intelligence, by this reasoning, we should consider it unintelligent. However, is such understanding correct? The third and last aspect questions whether the Turing test is beneficial for an AI designed for specific tasks because the results do not bring any new data and conclusions.

Keywords: deep learning (DL); artificial intelligence (AI); Turing test; intelligence; learning; training data; language

Úvod

Z filozofického hľadiska sa umelá inteligencia (ďalej len UI) vníma ako systém, ktorý by mal simulovať alebo imitovať ľudskú inteligenciu. S predstavou o inteligentných počítačoch prišiel už v roku 1950 Alan Turing. Zároveň ponúkol test (imitačnú hru), prostredníctvom ktorého by malo byť možné odlišiť strojovú entitu od ľudskej. Úlohou stroja v tomto teste je práve napodobiť človeka do takej miery, aby vyšetrovateľ (alebo aj porotcovia v novších verziách testu) nemohol odhaliť, že komunikuje so strojom. S príchodom hlbokého učenia¹ (ďalej len DL) sa situácia mení, nakoľko sa tieto umelé systémy primárne nezameriavajú na simulovanie ľudskej inteligencie, ale na riešenie konkrétnych špecifických problémov bez nutnosti využívať ľudské postupy. Na druhú stranu sa DL dokázalo v určitých činnostiach nielen vyrovnáť človeku, ale ho i prekonať (napr. v hre GO). Z dôvodu, že cieľom takejto UI nie je nutne simulovať ľudskú inteligenciu, sa otvára problém, či Turingov test (ako test pre UI napodobujúcu ľudskú inteligenciu) nie je v problematike hlbokého učenia irelevantný.

Prvá časť práce (2) sa bude zameriavať na aplikačné využitie Turingovho testu v tzv. Loebnerovej cene² (súťaž) v spojitosti s DL. Počas tejto súťaže sa pri umelých systémoch zisťuje absencia alebo prítomnosť ľudských aspektov inteligencie, ktoré zahŕňajú kategórie učenie, usudzovanie a porozumenie. V súvislosti s DL sa možno v kategórii učenia zamerať na metódu učenie s posilňovaním, u ktorej sa najpreukázateľnejšie objavujú určité prvky učenia. Dôležitým bodom pri riešení tohto problému je zistiť, či je možné uvádzanú metódu strojového učenia (ďalej len ML) považovať za použiteľný (i keď nebiologický) model učenia, a to v komparácii s preddefinovanou UI nedisponujúcou žiadnym typom ML, ale i s ľudským učebným procesom. V kategórii usudzovania sa analýza bude sústreďovať na tri druhy usudzovania – dedukcia, indukcia, abdukcia. Hlavným problémom bude v tomto prípade to, či sú tieto druhy v DL aplikované a akým spôsobom i v súvislosti s človekom. Porozumenie je aj u DL problematickým aspektom, nakoľko nie je možné jasne identifikovať, čo to znamená, že „UI rozumie“. Napriek tomu sa nevyklučuje možnosť, že DL disponuje určitým stupňom (alebo formou) porozumenia, vzhľadom na to, že nie je preddefinovaným, ale učiacim sa systémom, i keď forma porozumenia nemusí striktné korešpondovať s ľudskou predstavou tohto aspektu. Na základe uvádzaných troch kategórií sa objavuje možnosť irelevantnosti Turingovho testu v problematike hlbokého učenia, a to z dôvodu, že tieto kategórie nemusia byť striktné chápané iba v spojitosti s ľudskou inteligenciou, ale i s inými umelými systémami využívajúcimi odlišné postupy než človek. Rozoberanú tézu zároveň môžu podporiť aj dve nasledujúce časti, ktoré celý problém dopĺňajú.

Druhá časť (3) sa bude venovať otázke, či je v spojitosti s Turingovým testom možné považovať DL za neinteligentné. V Turingovom teste sa predpokladá, že UI bude simulovať všeobecnú ľudskú inteligenciu, nakoľko sa má čo najviac pripodobniť človeku, aby úspešne absolvovala túto skúšku. Keďže sa DL nepokúša a ani nedisponuje schopnosťami, ktorými by sa vyrovnala ľudskej inteligencii, je možné považovať tieto umelé systémy automaticky za neinteligentné, alebo test chápe inteligenciu nesprávnym spôsobom?

Tretia časť (4) sa zameria na pozíciu umelých systémov v Turingovom teste. Preddefinované chatboty sú priamo zamerané na absolvovanie takýchto testov, nakoľko majú naprogramovaný

¹ Hlboké učenie (deep learning, DL) založené na viacvrstvových neurónových sieťach je podkategóriou strojového učenia (machine learning, ML), ktoré je jedným z odvetví v umelej inteligencii (UI) (pozri Chollet 2019).

² Loebnerova cena je každoročná súťaž umelej inteligencie zameraná na konverzáciu v ľudskom jazyku. Založená bola v roku 1990 Hughom Loebnerom, po ktorom je aj pomenovaná. Súťaž prebieha na základe „štandardného Turingovho testu“ (Turing 1950).

bežný repertoár ľudských výpovedí. DL so zameraním na tréning jazyka môže mať vzhľadom na množstvo natrénovaných dát (a prepracovanosť algoritmu) preto výhodu v porovnaní so štandardným chatbotom. Na druhú stranu je otázne, aký zmysel by malo, aby takáto UI absolvovala Turingov test, keďže sa primárne zameriava na tréning štrukturálnych aspektov jazyka, na ktoré samotný test odpovede neprináša.

Turingov test v súčasnej umelej inteligencii

Alan Turing vo svojom článku *Počítacie stroje a inteligencia* predstavuje tézu o mysliacich strojoch, ktoré sa svojou inteligenciou vyrovnajú bežnému racionálnemu človeku. Turing sa však v práci nezaobrá analyzou ľudskej inteligencie a ani inteligenciou všeobecne. Namiesto toho ponúka test, prostredníctvom ktorého by malo byť možné odlišiť strojovú entitu od ľudskej. Pôvodne bol tento test (tzv. imitačná hra) považovaný skôr za myšlienkový experiment než za reálnu skúšku strojovej inteligencie. A to prevažne z toho dôvodu, že v dobe, v ktorej Turing žil, bol vrcholom technológie mechanický výpočtový stroj, a nie digitálny počítač alebo UI. Zásadná zmena prišla so založením Loebnerovej ceny – súťaže, ktorá Turingov experiment aplikovala ako plnohodnotný test. Na základe neho je možné odhaliť prítomnosť alebo absenciu ľudských aspektov inteligencie u tzv. chatbotov, čiže UI zameranej na konverzáciu v ľudskom jazyku.

Vývoj UI sa ale postupne začína zameriavať na pokročilejšie druhy tejto technológie, ktoré odsúvajú záujem o chatboty do vedľajšej pozície, keďže sú naprogramované iba na „hranie“ v Loebnerovej cene bez žiadneho ďalšieho účelu. Dôležitým míľnikom a zároveň jedným z hlavných smerov vývoja UI sa v súčasnosti stáva technológia hlbokého učenia, ktorého metódy a ciele sa diametrálne líšia od štandardných chatbotov. DL sa primárne nepokúša simulovať ľudskú inteligenciu, naopak je účelovo zamerané a natrénované na jednu konkrétnu úlohu (napr. AlphaGO na hranie hry GO). Dokonca je nutné dodať, že v súčasnosti ani nie je plnohodnotná simulácia možná, nakoľko sa vývoj ani nepribližuje k úplnému porozumeniu aspektov ľudskej inteligencie. Odborník na umelú inteligenciu M. Jordan v rozhovore s L. Gomesom dokonca tvrdí, že: „neurovedy sú ďaleko od riešenia, akým spôsobom sa ľudský mozog učí, tým pádom ani metódy hlbokého učenia nemôžu mať žiadnu priamu spojitosť s biologickým učením“ (Gomes 2014).

Účelová umelá inteligencia

Už v roku 1998 Ford a Hayes predpokladali, že lepší prístup k vývoju UI bude vtedy, keď „opustíme cieľ vytvoriť program imitujúci ľudskú inteligenciu a namiesto toho sa zameriame na systémy nevyžadujúce podobný prístup“ (Ford – Hayes 1998, 80). Vzhľadom na súčasný stav umelej inteligencie je možné tento predpoklad považovať za správny. Zároveň sa ukazuje, že existujú i odlišné prístupy k riešeniu UI, ktoré sú efektívne aplikovateľné aj v súčasnosti. Z toho dôvodu by mala aj filozofia zaujať nové stanovisko k tejto problematike. To znamená, že okrem neustáleho argumentovania a odôvodnenia možnosti „umelej ľudskej inteligencie“ (humanlike) je potrebné prijať aj novú kategóriu „umelej účelovej inteligencie“, do ktorej by bolo možné zaradiť aj DL, nakoľko sa primárne nesnaží simulovať biologický model učenia. Práve naopak, prichádza s novými metódami, ktoré sa so spomínaným modelom priamo nespájajú. V súvislosti s navrhnutým stanoviskom je nutné prehodnotiť relevanciu Turingovho testu v súčasnej UI, nakoľko ide z hľadiska filozofie mysle o prvú teóriu, ktorá otvorila problematiku mysliacich strojov. Keďže sa test postupne vytráca zo súčasného filozofického diskurzu, zostáva už iba otázka, či má miesto v aktuálnych problémoch UI, alebo je i pre DL bezvýznamný. Konkrétne sa v nasledujúcich pasážach budem zaoberať atribútmi inteligencie zisťovanými

prostredníctvom Turingovho testu, na základe ktorých sa pokúsim poukázať, že v metódach DL zohrávajú značne inú úlohu v porovnaní s ľudskou inteligenciou. Prípadne, že ich analýza pomocou tejto technológie ani nevyžaduje. Tým sa zároveň budem snažiť poukázať na možnosť, že Turingov test je v problematike hlbokého učenia irelevantný.

Predtým než budem pokračovať, by som rád zmienil niekoľko krátkych pripomienok. Nemám záujem v tejto práci poukazovať na všeobecnú bezvýznamnosť Turingovho testu, nakoľko je možné, že v budúcnosti bude existovať UI, pre ktorú sa stane vhodnou skúškou inteligencie (napr. UI so štruktúrou identickou s ľudským mozgom). Na druhú stranu sa chcem vyhnúť špekuláciám, ktoré nekorešpondujú s aktuálnym výskumom. Z tohto dôvodu aj (aktuálne) navrhujem pragmatickejšie stanovisko účelovej umelej inteligencie.

Učenie, usudzovanie a porozumenie v hlbokom učení

V Loebnerovej cene sa počas Turingovho testu zisťuje prítomnosť alebo absencia aspektov inteligencie (prostredníctvom kladených otázok), ktoré zahŕňajú učenie, usudzovanie a porozumenie. Na ich základe by malo byť možné rozlíšiť ľudskú entitu od strojovej. James Moor argumentuje, že „práve tieto kategórie sú perfektne spojené s ľudskou inteligenciou“. Zároveň tvrdí, že „nejde o nezávislé kategórie, ale práve naopak navzájom prepojené“ (Moor 2001, 87–88). S druhou časťou argumentu je možné súhlasiť, a to aj v súvislosti s DL. V prípade tejto UI ide dokonca o hierarchické prepojenie, nakoľko usudzovanie a porozumenie sú závislé od procesu učenia. Prvá časť argumentu je problematickejšia. Moor vo svojej práci nepredpokladá dve (aktuálne) kľúčové možnosti, ktoré by sa v rámci problematiky hlbokého učenia mohli objaviť. Po prvé, neberie v úvahu existenciu UI, ktorá by žiadnym spôsobom nesimulovala a ani nenapodobňovala ľudskú inteligenciu. Po druhé, považuje uvedené kategórie za striktné aspekty ľudskej inteligencie, čo je v súčasnosti možné označiť minimálne za nepresné hľadisko. Na objasnenie tohto problému sa je nutné podrobnejšie pozrieť na konkrétne kategórie, a to v rámci metód, ktoré DL využíva pri riešení úloh.

Kategória: učenie

Kľúčovou metódou, ktorú DL využíva pri riešení úloh, je tzv. učenie s posilňovaním.³ Prostredníctvom tejto metódy je schopné počas riešenia problému vykonať tri dôležité kroky: natrénovať surové dáta, samostatne optimalizovať dáta (UI si je v podstate sama sebe súperom, napr. AlphaGO hrala proti sebe hru GO) a nakoniec spätne si opraviť chyby (spätne šírenie chyby). Yamins a DiCarlo podotýkajú, že „takýto učiaci proces sa zdá byť bytostne odlišný od všetkého, čo je spojené s biologickým vývojom alebo ľudskou kultúrou“ (Yamins – DiCarlo 2016, 364). Na prvý pohľad sa zdá, že ide o relevantný názor. Najmä z dôvodu, že DL vykonáva učenie v uzavretom prostredí zameranom na jeden konkrétny problém. Inými slovami UI je (nebiologicky) kognitívne uzavretá⁴ v riešení špecifického problému v súlade s natrénovanými dátami, čo je diametrálne odlišné od ľudského spôsobu riešenia problémov. Overenie hypotézy, že metóda učenie s posilňovaním je skutočne odlišná od biologického modelu učenia, vyžaduje dva prístupy. Po prvé je nutné konfrontovať postupy DL so štandardnou UI, ktorá nevyužíva žiadne princípy ML. Na základe toho sa dá potvrdiť, že je zmieňovaná metóda reálne použiteľným modelom učenia. Druhým krokom je jej porovnanie s ľudskými postupmi, na základe čoho sa preukáže podobnosť alebo odlišnosť DL, a ľudskej inteligencie v kategórii „učenie“.

³ Okrem učenia s posilňovaním (reinforcement learning) sa medzi základné metódy strojového učenia (vrátane DL) ešte zaraďuje učenie s učiteľom (supervised learning) a učenie bez učiteľa (unsupervised learning).

⁴ Okrem prípadov, kedy sa využíva tzv. učenia s prenosom (transfer learning), prostredníctvom ktorého je možné natrénovanú neurónovú sieť použiť aj na riešenie súvisiacich podobných problémov (pozri Tan et al. 2018).

Predchodcovia DL, ktorí nevyužívali žiadnu formu ML, disponovali iba preddefinovaným naprogramovaním. To síce znamenalo, že takáto UI bola schopná vykonať úlohu, na ktorú bola nastavaná, no na druhú stranu u nej absentovala možnosť samostatnosti či učenie sa z reprezentácie dát,⁵ ktorá by prispela k podrobnejšiemu riešeniu určitého problému. Išlo vlastne o program, ktorého algoritmy nijakým spôsobom neaplikujú učebné prvky, a to i v prípade, že daná UI zdanlivo zvládala úlohy na ľudskej úrovni. Ukázkovým príkladom takéhoto druhu UI je Deep Blue, ktorý síce dokázal poraziť v šachovom dueli majstra sveta Garyho Kasparova, avšak iba na základe matematicky najpravdepodobnejších ťahov. Skutočnosť, že Deep Blue nedisponuje žiadnymi učebnými prvkami, potvrdzuje aj Schubbach: „Funkčnosť Deep Blue bola definovaná formálnymi algoritmami v programovacom jazyku C“ (Schubbach 2019, 1810), ktorý sa práve používa na preddefinované kódovanie. Zároveň sa ukazuje, že cieľom tejto UI bolo iba ukázať, ako technológia tej doby a preddefinované kódovanie dokáže poraziť človeka, avšak bez potreby vytvárať niečo nové. Dokonca aj Hsu, jeden z programátorov Deep Blue, hovorí: „Vedel som presne aké hardvérové hodnotenie bolo nastavené v každej hre“ (Hsu 2004, 200). Týmto tvrdením vlastne potvrdzuje preddefinovanosť krokov, ktoré Deep Blue vykonával počas hry. Na základe zmienených dôvodov možno usúdiť, že tento druh UI neprináša pri analýze nič nové okrem toho, čo už do nej vložili tvorcovia. V takomto prípade nie je možné hovoriť o tom, že by podobná UI disponovala určitou formou modelu učenia, nakoľko sa ani v jej postupoch neobjavujú znaky, ktoré by tomuto atribútu zodpovedali.

Učenie v hlbokom učení

Výhodou DL na rozdiel od predchádzajúceho variantu je už zmieňovaná metóda učenia s posilňovaním, ktorá ju oslobodzuje od nutnosti disponovať preddefinovaným naprogramovaním. Uvádzané tri kroky (natrénovanie dát, samostatná optimalizácia dát a spätná oprava chýb), ktoré sú integrálnou súčasťou metódy, zároveň poukazujú na možnú samostatnosť a učebné prvky v procese tréningu. Otázkou ostáva, či je takýto postup možné považovať za použiteľný (i keď nebiologický) model učenia. Relevantným zástupcom UI s DL je práve AlphaGO,⁶ na základe ktorej bude možné preukázať, či je metóda učenie s posilňovaním reálne použiteľným modelom učenia. Keďže zmieňované tri kroky sú v metóde kľúčové, rozbor problému bude pokračovať bližším pohľadom priamo na ne.

V prvom kroku AlphaGO natrénovala značné množstvo dát, ktoré obsahovali pravidlá a postupy hry GO, a čo je dôležité, duely odohrané ľudskými hráčmi (dáta približne z 1000 hier). Hneď tu sa je potrebné zastaviť a všimnúť si zásadné odlišnosti v porovnaní s Deep Blue. AlphaGO sa na rozdiel od Deep Blue prispôbuje počas natréningu hernému prostrediu, nejde preto primárne iba o využitie matematicky najpravdepodobnejšieho ťahu, ale skôr využitie dát ako „skúsenosti“ k ďalšej aplikácii. López-Rubio uvádza, že ide vlastne o „adaptáciu dát

⁵ Spôsob, akým sa v neurónovej sieti nahliada na dáta (ako ich DL vidí). Obrázky môžu byť reprezentované napr. v pixeloch alebo RGB. Ide o jeden z dôležitých algoritmov DL, na ktorom závisí správnosť výstupu (pozri Bengio et al 2013).

⁶ Z technického hľadiska je ešte potrebné upresniť, že AlphaGO mala okrem algoritmu učenie s posilňovaním implementované aj ďalšie algoritmy – učenie s učiteľom, vyhľadávací algoritmus Monte Carlo, hodnotiacu sieť (value network) a ešte niekoľko ďalších. To znamená, že jednotlivé kroky a celkovo celý proces natréningu nezávisel iba na jednej metóde, ale na kombinovanom použití algoritmov (pozri Silver, D. – Huang, A. – Maddison, C. et al. 2016). Zároveň ale možno dodať, že učenie s posilňovaním je jednou s kľúčových metód, na ktoré sa zameriava výskumný program od Google „DeepMind“ zaoberajúci sa vývojom DL, ako je AlphaGO, AlphaGO-Zero či AlphaStar.

k čo najlepšej sumarizácii problému“ (López-Rubio 2018, 671). Netreba ale zabúdať, že pojem skúsenosť sa v tomto prípade musí chápať striktnie v spojení s dátami, ktoré má DL k dispozícii. Najmä z dôvodu, že AlphaGO je (nebiologicky) kognitívne uzavretá v pravidlách a postupoch hry z natrénovaných dát, v ktorých síce vyniká, avšak nezrovnalosti nekorešpondujúce s hernými mechanizmami sú pre ňu neriešiteľné (pozri Hriadel 2020). I keď AlphaGO podlieha určitým obmedzeniam, v porovnaní s Deep Blue ide o nezanedbateľný pokrok, nakoľko je schopná použiteľným spôsobom zosumarizovať dáta, čo je zároveň u nej iba začiatkom riešenia, a nie finálnym stavom.

Ak by bolo možné nazvať prvý krok jedným slovom „sumarizácia“, druhý by sa bez problémov dal označiť ako „optimalizácia“, keďže po natrénovaní dát ich bola AlphaGO schopná plnohodnotne využiť. To sa prejavilo tým, že si samostatne nasimulovala herné duely, v ktorých hrala proti sebe samej. Zároveň zužitkovala nadobudnuté znalosti herných kombinácií, na základe ktorých vytvorila vlastné potenciálne výherné stratégie. V tomto bode sa ukazuje, že DL je schopné nadviazať a asociovať predchádzajúce znalosti a tým si vytvoriť vlastné poznatky, ktoré neboli vopred naprogramované. Niektorí by síce mohli povedať, že tieto schopnosti jej boli dané vďaka algoritmu, to by bol však nepresný názor, pretože algoritmus jej slúžil skôr ako šablóna podobne ako aj človeku. Treba aj dodať, že takýto prístup k riešeniu úlohy je unikátnym práve pre hlboké učenie. López-Rubio túto skutočnosť výborne konštatuje: „Automatické učenie je revolučným atribútom hlbokého učenia, nakoľko si takáto UI dokáže samostatne vybrať optimálnu možnosť.“ (López-Rubio 2018, 673) Tým sa taktiež snaží naznačiť, že v jednoduchších typoch strojového učenia (napr. plytkom učení⁷) táto vlastnosť absentuje (samozrejme to isté platí aj pre UI akou je Deep Blue). Zároveň tým poukazuje na samostatnosť, ktorá sa objavuje výhradne v algoritmoch DL.

Tretí krok – spätné šírenie chyby⁸ – by sa bez problémov mohol spojiť s predchádzajúcim, nakoľko sú vykonávané v podstate súčasne. I keď by sa na prvý pohľad mohlo zdať, že ide už iba o dodatkovú vlastnosť, nie je to tak, pretože je dôležitá pri finálnom konštruovaní stratégie. Presnejšie povedané, AlphaGO prostredníctvom tohto algoritmu znížila chybovosť vo svojich postupoch (konkrétne úpravou parametrov siete – váh), čo ju nakoniec doviedlo k výhernej stratégii, ktorou porazila v roku 2016 majstra sveta v hre GO Lee Sedola. Pravdepodobne aj preto ju Fitch nazýva „srdcom hlbokého učenia“ (Fitch 2014). Nakoľko je spätné šírenie chyby priamo prepojené s druhým krokom (samostatná optimalizácia dát), taktiež ide o výsadnú vlastnosť hlbokého učenia. Ako sa ukázalo, výsledok celého procesu natrénovania možno pokladať za značný úspech, nakoľko AlphaGO vytvorila ťahovú stratégiu, ktorá nebola nikdy vykonaná človekom. Tým sa preukázalo, že aj pomocou umelých metód učenia je možné prekonať človeka v činnostiach, ktoré majú k tomu ešte navyše aj veľmi dlhú tradíciu (hra GO vznikla približne v roku 2000 pred n. l.).

Na základe uvedených krokov je zjavné, že metóda učenie s posilňovaním prináša postup riešenia problémov s učebnými prvkami, ktorý je možné považovať za reálny a použiteľný model učenia, a nie iba za preddefinované naprogramovanie. Relevantnosť tézy podporuje aj skutočnosť, že preukázaný model učenia prináša výsledky, na základe ktorých je DL nielen schopné poraziť človeka v určitých činnostiach, ale dokonca vytvoriť aj niečo nové a neočakávané. Po predstavení vyššie spomínaného stanoviska treba konštatovať, že aj v technológii hlbokého učenia sa objavuje použiteľný model učenia. Vystáva ešte otázka, či je tento model učenia skutočne odlišný od ľudského, alebo sa v ňom nachádzajú určité podobnosti.

⁷ Shallow learning.

⁸ Spätné šírenie chyby (Error backpropagation alebo iba backpropagation) je z technického hľadiska algoritmus, ktorého úlohou je v neuronovej sieti znížiť chybovosť z predchádzajúcich vrstiev.

Komparácia s človekom

Levesque argumentuje, že sú dva hlavne spôsoby, akými sa človek môže učiť. Po prvé ide o „učenie zo skúsenosti, ktoré spája aj s tréningovým procesom“. V druhom prípade hovorí o „učení prostredníctvom jazyka, ktoré zase spája s kontextom a naratívnom“ (Levesque 2017). Na prvý pohľad by sa mohlo zdať, že sa v prvom prípade objavujú náznaky podobnosti. Druhý postup je naopak v súvislosti s UI (aj DL) problematický. Na základe predložených východiskových bodov sa je teraz nutné pozrieť na problém podrobnejšie z pohľadu oboch spôsobov a zároveň v spojitosti s postupmi DL, ktoré boli rozoberané v predchádzajúcich pasážach.

Učenie zo skúsenosti je s človekom späté od útleho detstva, keďže mu pre nadobudnutie znalostí touto cestou postačujú zmysly a interakcia s vonkajším svetom. Na rozdiel od prvej skúsenosti učenie prostredníctvom tréningového procesu vyžaduje k nadobudnutiu určitej zručnosti opakovaný prístup. Zároveň však každý potvrdí, že ide o veľmi prirodzenú cestu, ako sa niečo naučiť. Akákoľvek UI (vrátane DL) primárne nedisponuje hardvérom, ktorý by slúžil k adaptácii na vonkajšie prostredie. Dokonca v prípade DL je takýto doplnok úplne irelevantný, nakoľko sú jej algoritmy zamerané na učenie dát špecifického problému, a nie na interakciu s vonkajším svetom.⁹ To principiálne vylučuje podobnosť DL modelu učenia s ľudským.

Zdanlivo najvýraznejšia podobnosť sa objavuje v učení na základe tréningového procesu. Je možné spomenúť niekoľko kľúčových rozdielov, na základe ktorých bude možné preukázať, že v skutočnosti je táto domnienka nepresná alebo nepravdivá. Po prvé, i keď sa v oboch prístupoch učenia (u človeka a DL) objavuje pojem skúsenosť, jej použitie je iné. Človek nevyužíva enormné množstvo dát, aby sa zdokonalil v určitej zručnosti, namiesto toho spolieha na opakované pokusy, prostredníctvom ktorých sa postupne zlepšuje. DL naopak na sumarizáciu určitého problému (čiže vytvorenie „skúsenosti“ využitej na riešenie úlohy) vyžaduje veľké množstvo dát, inak by nebola schopná vykonávať ďalšie kroky.¹⁰ Po druhé, DL je schopné vykonávať spätnú opravu chýb v celej neurónovej sieti, aby minimalizovala chybovosť. V tomto bode je nutné povedať, že aj človek nepochybne vie spätne opraviť svoje chyby, z ktorých sa poučí. Nie je však striktné viazaný na naučené dáta, v ktorých by hľadal potenciálne nezrovnalosti. Tréningový proces u človeka je skôr postavený na pokusoch a omyloch. Fitch túto diferenciu trefne komentuje: „Spätná oprava chýb je biologicky plauzibilná, nie však na úrovni neurónovej siete“ (Fitch 2014, 340). Po tretie, človek sa ani nepribližuje úrovni kognitívnej kapacity, ktorá by mu umožňovala samostatne vytvárať riešenia (napr. stratégie), pri ktorých by si mohol byť simultánne aj oponentom, tak ako je schopná DL (tým zároveň nechcem povedať, že by si človek v žiadnom prípade nemohol byť oponentom). V súvislosti s nastolenými rozdielmi sa nakoniec ukazuje, že zdanlivá podobnosť je mylným názorom. Je zjavné, že sa tréningové procesy u človeka a DL neporovnateľne líšia. Dokonca je možné povedať, že sú to dva odlišné učebné postupy.

Učenie prostredníctvom jazyka je aj v súčasnosti nezlučiteľné s postupmi UI, a to i v prípade, že je možné hovoriť o použiteľnom modeli učenia, keďže v prípade DL sú vstupné dáta

⁹ Tým nechcem povedať, že tréňované dáta nemôžu byť o vonkajšom svete. Práve naopak. Dáta sa vzťahujú k riešenému problému (napr. dáta v podobe cestných záznamov, jazykové nahrávky, texty či dáta o hre). Samotné natreňovanie však prebieha na počítačoch s výkonným hardvérom s mnohými GPU a CPU. Integrácia (čiže konkrétne použitie) napr. u áut prebieha až po natreňovaní, a nie v aute.

¹⁰ Výnimkou je jednorazové učenie (one shot learning), ktoré vyžaduje menšie množstvo dát (pozri Vinyals et al. 2016). I keď stále nejde o konceptuálne učenie podobné človeku, tento prístup je tomu najbližšie. Na druhú stranu zatiaľ neprináša zatiaľ také výsledky (aj aplikácie využité) ako DL využívajúce veľké databázy.

natrénované štatistickým spôsobom. U ľudí je samozrejme situácia úplne iná, či už ide o prijímanie znalostí z kontextového porozumenia, alebo prostredníctvom naratívu. Možnosť využívať sémantiku umožňuje človeku priame učenie bez nutnosti vykonať hlbšiu analýzu, a to iba prostredníctvom veľmi malého počtu informácií (napr. jedno slovo). Goulden a jeho kolegovia napr. uvádzajú, že „dieťa sa je schopné naučiť novú znalosť už na základe niekoľkých slov a fráz“ (Goulden et al. 1990). Ako som už uvádzal, DL by v takejto situácii nebolo schopné vykonávať natrénovanie, keďže k tomu vyžaduje enormné množstvo dát. Nechcem tu hovoriť o tom, že keď UI nedisponuje sémantikou, nemôže ani myslieť. Zároveň nesúhlasím, aby sa inteligencia striktne spájala s myslením, keďže sa ukázalo, že môže fungovať použiteľný model učenia aj bez ľudských atribútov. Namiesto toho navrhujem, aby sa deficit sémantiky nepovažoval nutne za nevýhodu, na základe ktorej mnohí ešte aj dnes označujú UI za neužitočný nástroj. Ide len o iný prístup riešenia problémov využívajúci štatistické metódy namiesto sémantického prístupu, ktorý však prináša výsledky bez potreby vyniknúť v myšlienkovom experimente (argument čínskej izby) (pozri Searle 1994). Ide však o úplne iný problém, ktorý priamo nekorešponduje s aktuálnou témou, preto ho nateraz nechávam otvorený.

Zdá sa, že odlišnosti medzi ľudským a umelým modelom učenia sú natoľko zjavné, že oba modely je nutné považovať za samostatne použiteľné. Preukázalo sa, že umelý model učenia implikuje učebné prvky a postupy prinášajúce nepopierateľné výsledky. Niet pochýb, že DL spĺňa kategóriu „učenie“ bez nutnosti integrácie alebo simulovania ľudských aspektov učenia, ktoré sú diametrálne odlišné od postupov prostredníctvom metódy učenia s posilňovaním.

Kategória: usudzovanie

Problematike usudzovania v DL (ale i všeobecne v ML) sa nevenuje toľko pozornosti ako učeniu (trénovaniu dát). Dôvodom môže byť to, že tvorcovia DL sa zväčša zaoberajú iba algoritmami či neurónovými sieťami, ostatné aspekty však opomínajú. Ako aj upozorňuje Zhou, „strojové učenie a logické uvažovanie sú v histórii UI skúmané oddelene“ (Zhou 2019, 1). Napriek oddelenému výskumu sa niekoľko počítačových vedcov zaoberá aplikáciou indukcie a abdukcie v ML (Zhou 2019; Kakas – Michael 2020; Mooney 2000; Nguifo 1998). Na základe toho sa pokúsim v tejto časti poukázať na opodstatnenosť týchto druhov usudzovania v DL. Zároveň rozoberiem, prečo sa tretí druh usudzovania – dedukcia – s postupmi DL nespája.

Dedukcia

Z psychologického hľadiska ľudia využívajú všetky druhy usudzovania intuitívne. Dokonca nemusia ani vedieť, ktoré z nich v danú chvíľu použili (ak neberieme v úvahu výskum v logike). Na rozdiel od človeka DL pracuje prostredníctvom presných komputačných postupov, ktoré úplne ľudskú intuíciu nezahŕňujú. Jednou z úloh, ktorú DL vykonáva počas tréovania, je zovšeobecnenie kontextu, z ktorého môže následne predikovať nové dáta. Keďže je výsledkom detektívneho usudzovania opak a zároveň DL prijíma prvotné dáta, ktoré nie sú zovšeobecnené, dedukcia by v takomto kroku nemala žiadne využitie. Nadväzujúcim problémom je, že dedukcia neprináša novú vedomosť. Oterllo konkrétne uvádza, že „dedukcia vedomosti odvodzuje, indukcia a abdukcia môžu priniesť nové poznatky“ (Oterllo 2013, 56). Pokiaľ je úlohou DL predikovať či vytvárať nové dáta (a tým aj poznatky), dedukcia ani túto podmienku nespĺňa. Zároveň sa tým objasňuje, prečo sa mnohí zameriavajú na výskum indukcie a abdukcie v ML, a nie na tento druh usudzovania.

Indukcia

Za predpokladu, že indukcia a abdukcia sú súčasťou postupov ML, treba objasniť ich konkrétnu úlohu v procese natrénovania. Podľa Facha a Kakasa „nededuktívne druhy usudzovania vykonávajú v UI dva základné procesy – explanáciu a zovšeobecňovanie“ (Flach – Kakas 2000, 15). Ako už bolo naznačené, jedným z postupov DL je práve zovšeobecniť tréňované dáta. Presnejšie je to možné rozvíeť tak, že čiastkovou úlohou v DL je v prijatých dátach nájsť relevantné vzorky informácií, z ktorých môže vytvoriť všeobecné okruhy, a z tých generovať nové výsledky, čiže dáta vzťahujúce sa k riešenému problému. Napriek tomu, že celý proces nezodpovedá indukčivnému uvažovaniu, je možné na základe Nguifovmu výkladu indukcie – „entita – fakt – overenie vlastnosti – fakt – zovšeobecnie faktov“ (Nguifo 1998, 2) – usúdiť, že určitá časť tréňovania je indukčivná.

Abdukcia

Abdukcia je proces vytvárania explanácii (hypotéz) z pozorovaných dát. V porovnaní s predchádzajúcimi druhmi usudzovania ide o menej známy logický proces. Na druhú stranu v problematike ML sa viacerí autori zameriavajú prevažne na abdukciu (autori spomínaní vyššie). Prečo je práve abduktivné usudzovanie dôležité pri tréňovaní dát, napovedá argument Kakasa a Michaela: „Abduktivné usudzovanie môže byť použité na vyplnenie medzier u nekompletných vzoriek dát prostredníctvom informatívnych hypotéz o ich hodnotách, čím pomáha pripraviť tréningové dáta na proces učenia“ (Kakas – Michael 2020, 10). Uvádzaný argument je kľúčový z dvoch dôvodov. Po prvé, implicitne poukazuje na skutočnosť, že abdukcia v ML predchádza indukčivii. To podporujú aj etapy vedeckého skúmania, v ktorých generovanie hypotéz (abdukcia) zohráva prvú úlohu (Fach – Kakas 2000, 6, 7). Po druhé, napovedá akú úlohu zohráva abduktivné usudzovanie v tréňovaní dát. Keďže sa prostredníctvom abdukcie vytvárajú explanáčnne hypotézy z pozorovaných (prijatých) dát, možno usúdiť, že v ML zastupujú základné vzorky znalostí, ktoré sa následne indukčivným usudzovaním zovšeobecňujú.

Ako naznačuje posledná veta, medzi indukčivou a abduktivou existuje i v ML určitý vzťah, ktorý možno rozdeliť do dvoch kontinuálne sa opakujúcich etáp. Za prvú etapu možno označiť už zmieňované prepojenie, pri ktorom sa z výstupu abduktivného usudzovania stáva vstup pre indukčivné, pomocou ktorého sa relevantné dáta (hypotézy) zovšeobecnia do nových znalostí. Druhá etapa môže nastať v prípade, ak sú dáta aktualizované, prípadne nekompletné. To znamená, že proces prebieha dovtedy, pokiaľ spracovanie dát nie je dostatočné. Flach a Kakas (2000) aj Kakas a Michael (2020) sa zhodujú, že ide o tzv. cyklus abduktivného a indukčivného rozvoja vedomostí.¹¹

Na základe analýzy troch druhov usudzovania v súvislosti s DL sa ukázalo, že procesy prostredníctvom dedukcie nekorešpondujú s postupmi počas tréňovania dát. Na druhú stranu sa niektoré kroky, ktoré DL vykonáva, ukázali ako indukčivné. Posledný druh usudzovania – abdukcia – zohráva taktiež dôležitú úlohu v procese tréňovania, nakoľko pripravuje základné vzorky znalosti, t. j. hypotézy, ktorá sa ďalej prostredníctvom indukcie zovšeobecňujú. Zároveň je potrebné skonštatovať, že medzi indukčivou a abduktivou existuje v DL vzťah nazývaný cyklus abduktivného a indukčivného rozvoja vedomostí. V súvislosti človekom je nakoniec možné uviesť, že v prípade DL nejde o intuitívne využívanie usudzovania, ale viazané na proces natrénovania dát.

¹¹ The cycle of abductive and inductive knowledge development.

Kategória: porozumenie

Kategóriu „porozumenia“ je i v súvislosti s DL možné označiť za funkcionalistickú čiernu skrinku (black box),¹² nakoľko nie je jasne identifikovateľné, čo vlastne znamená, že „UI rozumie“. Samozrejme, problém sa môže spájať s vyžadovaním explanácie. UI nie je schopná nijakým spôsobom vysvetliť, prečo práve vykonala ten určitý krok. Zároveň je nutné dodať, že by sa jej to nepodarilo spôsobom, ktorý by bol pre ľudí prijateľný a zrozumiteľný. Ľudia na druhú stranu vyžadujú explanáciu ako dôkaz, že tomu, čo hovoria, skutočne rozumejú, a to na základe významov. Vynára sa otázka, či aj v DL, ktoré na rozdiel od preddefinovaných programov disponuje preukázateľným modelom učenia, ale aj formami usudzovania, skutočne neexistuje schopnosť porozumenia alebo je diametrálne odlišná od ľudského variantu. Druhá možnosť sa zdá byť vzhľadom na predchádzajúce dve kategórie pravdepodobnejšia.

Existujú dozaista ľudia, ktorí bez problémov chápu komplexné matematické operácie a zároveň ovládajú niekoľko numerických sústav. Na základe toho dokážu vytvárať a analyzovať počítačové algoritmy. Nie je to však bežný spôsob, akým sa prijímajú či konštruujú nové informácie. Na rozdiel od UI sú ľudia uspôsobení na poznávanie prostredníctvom sémantického používania jazyka (a skúsenosti), čomu aj prispôbujú prostredie, v ktorom žijú, a to v globálnom hľadisku. Pre DL prostredie nepredstavuje celý svet plný významov, ale natrénovanú neurónovú sieť, na ktorú je primerane optimalizované a dokáže v nej vykonávať náležité kroky (na základe algoritmov). Ako uvádzajú Kruger a jeho kolegovia, „z každej vrstvy je vyextrahovaná vlastnosť, ktorá je adaptovaná na nasledujúcu vrstvu s komplexnejšou vyextrahovanou vlastnosťou, než mala predchádzajúca“ (Kruger et al. 2013, 1864, 1865). Na základe Krugeroých slov je možné usúdiť, že DL dokáže vytvárať v neurónovej sieti reprezentácie, s ktorými môže štruktúrne (vrstva nadväzuje na vrstvu) zaobchádzať. I keď ani Krugerova téza nedokazuje, že by bolo DL schopné rozumieť jednotlivým krokom, ktoré vykonáva. Minimálne upozorňuje na skutočnosť, že v neurónových sieťach sa vykonávajú komplexnejšie kognitívne operácie než v prípade klasickej UI.

Project Debater

Pravdepodobne jeden z najťažších kognitívnych úkonov a zároveň výziev vo výskume UI je komunikácia s človekom, a to nie len vo forme odpovedí na otázky, ale i plynulej diskusie na určitú tému. V tomto odvetví zatiaľ dosiahol najväčší úspech Project Debater od IBM, ktorý dokázal absolvovať v roku 2019 argumentačný duel proti ľudskému súperovi. Ako uvádzajú tvorcovia projektu, „Project Debater bol schopný prostredníctvom metód „ťažba argumentov“ (argument mining) určiť relevantné argumenty a protiargumenty v rámci všetkých natrénovaných textov, pomocou AKB vyselektovať najrelevantnejšie argumenty pre jednotlivé druhy diskusií, t. j. vybrať najvhodnejšie argumenty priamo na diskusiu. Na základe metódy „vyvrátenia argumentu“ (argument rebuttal) predpokladať tvrdenia oponenta a vytvoriť na ne odpoveď a nakoniec pomocou „konštrukcie diskusie“ (debate construction) vytvoriť argumentačnú reč“ (Slonim, N. – Bilu, Y. – Alzate, C. et al. 2021, 380). Project Debater síce nedokáže používať jazyk takou formou ako ľudia (sémantika), prostredníctvom uvádzaných metód sa však ukazuje, že existujú aj iné spôsoby, ako hodnotiť kvalitu a konštruovať argumenty (Toledo et al. 2019). V súvislosti s problémom porozumenia sa je možné na proces pozrieť z dvoch strán. Po prvé, keďže Project Debater nedisponuje rovnakými schopnosťami ani hardvérom ako človek, ktorý by umožňoval porozumieť argumentom a celému procesu diskusie, nie je schopný porozumenia. Po druhé, pokiaľ mu jednotlivé metódy umožňujú pracovať s textami, na základe

¹² Z funkcionalistického hľadiska je čierna skrinka (black box) myseľ (alebo UI program), u ktorej nie sú známe vnútorné mechanizmy, ale iba vstupné a výstupné informácie.

ktorých je schopný hodnotiť i vytvárať argumenty v rámci celej diskusie, nie je možné minimálne z funkcionalistického hľadiska určitú úroveň porozumenia vylúčiť.

Na rozdiel od predchádzajúcich dvoch kategórií (učenie a usudzovanie) je porozumenie v DL inherentnejším aspektom, ktorý nemá žiadne zjavné podobnosti s človekom. Vyžadovanie explanácie ako dôkaz porozumenia je bežnou záležitosťou v prípade sémantickej formy. V neuronovej sieti a súčasných jazykových systémoch sa však objavujú metódy, v ktorých by bolo možné určitý stupeň porozumenia predpokladať, alebo ho minimálne úplne nevyučujú.

Na základe prezentovaných zistení možno konštatovať, že existujú zásadné rozdiely medzi ľudskou inteligenciou a DL. Napriek uvedeným odlišnostiam je i tak zjavné, že DL disponuje atribútmi, ktoré mu umožňujú vykonávať proces tréningu dát bez nutnosti aplikovať prvky ľudskej inteligencie.

V súvislosti s analyzovanými kategóriami využívanými v Turingovom teste na odlíšenie strojovej a ľudskej entity je možné usúdiť, že ich síce DL spĺňa, avšak úplne iným spôsobom, ako sa predpokladá v tomto teste, keďže sa považujú za striktné ľudské aspekty, ktoré by mal adept na úspešné absolvovanie testu simulovať podľa človeka. Z účelového hľadiska je DL úplne odlišným systémom, ktorý prináša iné spôsoby riešenia problémov namiesto simulovania ľudskej inteligencie, a to i v rámci uvedených kategórií. Z tohto dôvodu je Turingov test v spojitosti s problematikou hlbokého učenia možné považovať minimálne za irelevantný.

Nízka kognitívna úroveň hlbokého učenia

V Turingovom teste sa predpokladá, že UI bude simulovať ľudskú inteligenciu, ktorú možno označiť ako všeobecnú. Inými slovami, človeku jeho inteligencia umožňuje nielen učiť sa z jazyka a skúsenosti, ale i utvárať spomienky a asociovať ich s novými myšlienkami, vyznávať určité náboženstvo, zaradiť sa do nejakého kultúrneho celku atď. Pokiaľ UI nie je schopná reagovať a odpovedať na otázky z týchto alebo podobných oblastí, je označená za neinteligentnú. French na odhalenie nízkej kognitívnej úrovne u UI (neschopnosť sa vyrovnáť ľudskej inteligencii) vytvára niekoľko subkognitívnych otázok, ktoré podľa neho nedokáže zodpovedať žiadny umelý systém. Napr. „Je slovo Flugblogs vhodné použiť ako názov novej počítačovej spoločnosti? Je slovo Flugly vhodné na pomenovanie detského plyšového medvedíka?“¹³ (French 1990, 59). Pravdepodobne ani najlepšia UI zameraná na spracovanie prirodzeného jazyka by nepodala uspokojivú odpoveď porovnateľnú s človekom, nakoľko tie slová neexistujú v štandardnej slovnej zásobe. Človek je na druhú stranu schopný podobné výrazy asociovať so slovami, ktoré už má vo svojej slovnej zásobe, dokonca ich môže spojiť s nejakou šťastnou alebo smutnou spomienkou. Na základe toho nakoniec dokáže podať uspokojivú odpoveď, s ktorou by každý pravdepodobne súhlasil a považoval ju za „ľudskú“. Znamená to, že máme UI a konkrétne DL považovať za neinteligentné? Pravdepodobne nie.

Inteligencia v hlbokom učení?

Problém je v samotnom pojme „inteligencia“, ktorý je častokrát striktné a i nesprávne spájaný iba s ľudským myslením a celým jeho kognitívnym systémom. Parnas napr. tvrdí, že

¹³ French ďalej aplikuje subkognitívne otázky do tzv. hodnotiacej hry (rating game), v ktorej sa má hodnotiť vhodnosť neologizmov v otázkach od 0 (kompletne neprijateľné) do 10 (kompletne prijateľné). Človek prostredníctvom sémantického prepojenia (napr. flugly – ugly) alebo kultúrnej asociácie dokáže výrazy vzhľadom na položené otázky bez problémov zhodnotiť. Umelá inteligencia by s tým (bez zmieňovaných kognitívnych dispozícií) mohla mať problém (French 1990).

akákoľvek UI nemá s inteligenciou nič spoločné: „Inteligentná UI je iba ilúziou, vytvorenou z nesprávneho porozumenia významu pojmu inteligencia, ktorý môže mylne naznačovať, že UI ma blízko k ľudskej inteligencii.“ (Parnas 2017, 31) Práve na Parnasových slovách sa ukazuje, ako veľmi si ľudia privilegujú a privlastňujú atribút „byť inteligentný“. Zároveň sa s týmto hľadiskom spája problém, že ľudia chápu dispozíciu inteligencie binárne, bez možnosti uvažovať o iných formách inteligencie. Inak povedané, buď ňou nejaká entita disponuje, alebo nie. Súhlasím, že preddefinovaná UI skutočne nie je inteligentná, keďže ide skôr o program vykonávajúci presne nastavené kroky bez náznakov inteligencie. V prípade DL však s týmto názorom nemožno súhlasiť. V predchádzajúcich častiach sa totiž ukázalo, že tento typ UI je schopný vykonávať komplexné procesy zahŕňajúce učenie, usudzovanie i určitý stupeň porozumenia, ktoré je možné považovať za preukázateľné znaky inteligencie. Súčasné DL by bolo možné považovať za neinteligentné iba v prípade, že by sa pokúšalo pripodobňovať ľudskej inteligencii (presnejšie povedané, bolo by ho možné označiť za neinteligentné v porovnaní s človekom). O to sa však ani nesnaží. Namiesto toho je účelovo natrénované na riešenie konkrétnych problémov, v ktorých prináša preukázateľné výsledky. Z toho je možné usúdiť, že DL disponuje formou inteligencie, ktorá je dostatočná na riešenie problémov, na ktoré je primárne zameraná.

Ako sa ukázalo, v Turingovom teste je inteligencia chápaná nesprávnym spôsobom, nakoľko sa v ňom redukuje iba na ľudskú inteligenciu a neberie v úvahu ďalšie potenciálne formy. Keďže sa však v DL objavujú preukázateľné znaky inteligencie, niet pochýb, že táto UI určitou formou inteligencie disponuje. To podporuje aj skutočnosť, že použité procesy prinášajú náležité výsledky v rámci riešeného problému. Skutočnosť, že DL disponuje formou inteligencie, ktorá sa nijakým spôsobom nesnaží pripodobňovať ľudskej, stavia Turingov test opäť do pozície, v ktorej je pre túto UI pravdepodobne irelevantný.

Pozícia hlbokého učenia v Turingovom teste

Chatboty s preddefinovaným naprogramovaním pravdepodobne nebudú mať nikdy všestrannejšie využitie než len súťaženie v „testoch inteligencie“ založených na Turingovom teste. Iná situácia nastáva v prípade chatbotov založených na technológii hlbokého učenia, ktorá nie je viazaná na podobné naprogramovanie, ale naopak na natrénovanie dát. To im zreteľne prináša väčší potenciál aj v tréningu jazyka (presnejší názov pre túto UI znie DL zamerané na spracovanie prirodzeného jazyka), podobne ako u AlphyGO v hre GO. Takouto UI je napr. GPT-3, ktorý ešte nedosahuje ľudskú úroveň v otázke používania jazyka. Ako uvádza Lacker, zvláda „základy logiky (logických otázok) a je schopný využívať zdravý rozum (common sense)“ (Lacker 2020). Nemám záujem hovoriť o tom, ako blízko je ľudskej inteligencii, nakoľko by to aktuálne ani nemalo význam. Samotná UI je totiž iba v začiatkovej beta verzii. Zároveň tento problém priamo nekorešponduje s témou štúdie. Namiesto toho sa zameriam na problém, do akých pozícií to stavia túto UI (a i ďalšie podobné systémy zamerané na spracovanie prirodzeného jazyka), a to i v súvislosti s Turingovým testom.

Po prvé, vyvstáva otázka, či má zmysel, aby podobné UI absolvovali imitačnú hru, nakoľko aj Turing redukuje možnosť myslenia či inteligencie v umelých systémoch na „úspešné absolvovanie imitačnej hry“ (Turing 1950). Samozrejme, natrénované DL by malo v porovnaní s preddefinovanou UI väčšiu šancu na úspech, nakoľko by bolo schopné aktívne pristupovať a simultánne využívať veľké množstvo dát. Princíp by však zostal rovnaký, a to oklamať vypočítavajúceho vyšetrotavateľa, eventuálne aj porotu v novších variantoch Turingovho testu. Nijakým spôsobom by sa v tomto teste nereflektovali postupy alebo i odlišná forma inteligencie v rámci DL, nakoľko by sa predpokladalo, že testovaný systém musí simulovať ľudskú inteligenciu.

Výsledkom by bolo iba odhalenie alebo neodhalenie (v prípade úspechu) funkcionalistickej „čiernej skrinky“ bez žiadneho ďalšieho prínosu.

Na druhú stranu, účelové využitie a presná reprezentácia dát prostredníctvom metód DL nielenže prináša výsledky, umožňuje aj riešenia veľmi špecifických problémov, na ktoré človek nemá kognitívnu kapacitu. Plytké pripodobňovanie sa človeku je možno signifikantné v rámci myšlienkového testu imitačnej hry, v skutočnosti to však neprináša nič nové na rozdiel od riešenia rôznych úloh (vrátane jazykových) prostredníctvom postupov DL. French napríklad argumentuje, že „niektoré situácie (interakcia so svetom) vyžadujú zapojenie ľudských zmyslov“ (French 2000, 119). Tým upozorňuje na obmedzené kognitívne schopnosti UI. S týmto tvrdením možno súhlasiť, je ale nutné dodať, že podobný deficit by nebol schopný odhaliť ani Turingov test. Napriek kognitívnym obmedzeniam je DL schopné vykonávať postupy, na ktoré je určené, a to i v prípade spracovania prirodzeného jazyka. Samozrejme by neprinieslo a nebolo schopné adaptovať poznatky so situácií, s ktorými sa človek bežne stretáva. DL zamerané na spracovanie prirodzeného jazyka sa však sústreďuje na štrukturálnu stránku problému, a nie na ľudské skúsenosti.

Ak by bolo DL zamerané na spracovanie prirodzeného jazyka stávané do pozície, v ktorej by malo iba absolvovať Turingov test, neprinieslo by to nič nové. Záver testu by priniesol iba odhalenie či neodhalenie systému, ktorý by bol stále považovaný za „čiernu skrinku“ bez žiadnej ďalšej reflexie potenciálne odlišnej formy inteligencie. Z účelovej pozície však DL prichádza s rôznymi možnosťami využitia (komunikácia v ľudskom jazyku, hlasový asistenti, rozpoznanie reči atď.), a to bez nutnosti simulovať ľudskú inteligenciu v imitačnej hre, čo by i tak neprinieslo žiadny ďalší prínos okrem úspechu v tomto myšlienkovom experimente.

Záver

Turingov test je pravdepodobne vhodnou skúškou inteligencie pre systémy, ktoré sú naň priamo naprogramované. Na druhú stranu ho nie je možné ponímať ako všeobecný test pre akýkoľvek umelý systém, keďže práve v metódach súčasného DL sa objavujú aspekty, ktoré nemajú so simulovaním ľudskej inteligencie priamu súvislosť.

Tri kategórie inteligencie – učenie, usudzovanie a porozumenie – zisťované v aplikačnom využití Turingovho testu v rámci Loebnerovej ceny síce tieto umelé systémy spĺňajú (alebo nimi disponujú), avšak odlišným spôsobom než človek. DL oproti preddefinovaným systémom disponuje použiteľným (i keď nebiologickým) modelom učenia sa na základe metódy učenie s posilňovaním. V porovnaní s človekom má však diametrálne odlišný kognitívny (nebiologický) postup najmä s dôrazom na množstvo naučených informácií (v prípade UI dát). Kľúčové druhy usudzovania sú v DL abdukcia a indukcia. Prostredníctvom abdukcie sa z prijatých dát vytvárajú základné vzorky znalostí (hypotézy), ktoré sú následne zovšeobecnené pomocou induktívneho usudzovania. Zároveň medzi týmito dvoma druhmi usudzovania existuje vzťah, ktorý im umožňuje pri aktualizácii eventuálne nekompletnosti dát celý proces opakovať, pokiaľ nebude stav dát dostatočný na tréning. Dedukcia v tréningu dát nemá konkrétne využitie, nakoľko je dôležitým krokom v tomto procese zovšeobecňovanie, ktoré deduktívne uvažovanie neumožňuje. I keď sa v tomto prípade objavuje zjavná podobnosť s človekom, je potrebné zmieniť, že ľudia využívajú všetky druhy usudzovania intuitívne bez potreby špecificky analyzovať, ktorý druh usudzovania použiť. Využitie konkrétnych druhov usudzovania však v DL korešponduje s počítačnými metódami ML. Porozumenie v DL i tak nie je jasne identifikovateľné. Nie je ani možné s istotou povedať, že táto UI žiadnym spôsobom „nerozumie“. Aj keď nejde o porozumenie na základe sémantickej formy jazyka, ukázalo sa, že v prostredí a v metódach DL

prebiehajú procesy, ktoré schopnosť porozumenia úplne nevylučujú, i keď diametrálne odlišným spôsobom ako sa predpokladá u človeka. Analýza uvádzaných troch kategórií ukázala, že možnosť irelevantnosti Turingovho testu v problematike hlbokého učenia nie je vylúčená. Rozbor nasledujúcich dvoch problémov poukázal na aspekty, ktoré relevantnosť tézy ďalej podporujú.

Všeobecná inteligencia ľuďom umožňuje riešiť viacero rozličných problémov, ale i chápať svet a vlastný život spôsobmi, ktoré nie sú striktné viazané na naučené informácie. To však neznamená, že inteligencia musí byť ponímaná iba binárne. Inými slovami či ňou entita disponuje, alebo nie. Keďže je DL zamerané na riešenie konkrétnych problémov prostredníctvom metód ML s učebným procesom, nesmie byť bez pochybností označené za neinteligentné. Na rozdiel od človeka DL nepotrebuje (a ani nemá možnosť) pri tréningu brať do úvahy nuansy vonkajšieho sveta, nakoľko by aj k riešeniu konkrétneho problému mohli byť irelevantné. V súvislosti s metódami ML je možné inteligenciu v tejto UI ponímať ako inú úroveň alebo formu, ktorá je optimalizovaná na riešenie problémov, v ktorých prináša výsledky, a to aj na nadľudskej úrovni.

DL zamerané na spracovanie prirodzeného jazyka oproti chatbotom naprogramovaným k absolvovaniu Turingovho testu môže priniesť výsledky v podobe aplikačného použitia (integrácia do umelých systémov využívajúcich ľudský jazyk či reč). Za predpokladu, že by aj táto UI bola schopná absolvovať Turingov test (dokonca by mohla mať výhodu v súvislosti s natrénovanými dátami), neprinieslo by to žiadne ďalšie poznatky o možných formách alebo úrovniach inteligencie, okrem odhalenia či neodhalenia „čiernej skrinky“. Z tohto dôvodu pre účelovo ponímané DL, ktoré prináša výsledky v spracovávaní jazyka, nemá zmysel absolvovať Turingov test, nakoľko by to bol v konečnom dôsledku iba ďalší problém, na ktorý by muselo byť natrénované.

Na základe rozoberaných problémov je možné považovať Turingov test v súčasnej UI založenej na DL za irelevantný. Situácia by sa síce mohla zmeniť, keby neurovedy dokázali replikovať ľudský mozog a jeho vlastnosti. V súvislosti s aktuálnym vývojom UI je však irelevantnosť Turingovho testu v problematike hlbokého učenia preukázateľnou možnosťou.

PodĎakovanie:

Tento príspevok vznikol vďaka podpore APVV-17-0064 - *Analýza multidimenzionálnej podoby trans- a posthumanizmu*.

Bibliografia

Bengio, Y. (2013): Representation Learning: A Review and New Perspectives, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(8), 1–30, dostupné z: < 10.1109/TPAMI.2013.50 >.

Fitch, W. T. (2014): Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition, *Physics of Life Reviews* 11(3), 329–364, dostupné z: < <https://doi.org/10.1016/j.plrev.2014.04.005> >.

Flach, P. A. – Kakas, A. C. (2000): Abductive and Inductive Reasoning: Background and Issues, in Flach, P. A. – Kakas, A. C. (eds.) *Abduction and Induction. Applied Logic Series* 18, 1–27, dostupné z: < https://doi.org/10.1007/978-94-017-0606-3_1 >.

Ford, K. M. – Hayes, P. J. (1998): On Computational Wings: Rethinking the Goals of Artificial Intelligence, *Scientific American Presents* 9, 78–83.

French, R. M. (1990): Subcognition and the Limits of the Turing Test, *Mind* 99(393), 53–65, dostupné z: < <https://doi.org/10.1093/mind/XCIX.393.53> >.

French, R. M. (2000): The Turing Test: The First 50 Years, *Trends in Cognitive Sciences* 4(3), 115–122, dostupné z: < [https://doi.org/10.1016/S1364-6613\(00\)01453-4](https://doi.org/10.1016/S1364-6613(00)01453-4) >.

Gomes, L. (2014): Machine-Learning Maestro Michael Jordan on the Delusions of Big Data and Other Huge Engineering Efforts, *IEEE Spectrum* [online], 2014-10-20, [cit. 2021-03-31], dostupné z: < <https://spectrum.ieee.org/machinelearning-maestro-michael-jordan-on-the-delusions-of-big-data-and-other-huge-engineering-efforts> >.

Goulden, R. – Nation, P. – Read, J. (1990): How Large Can a Receptive Vocabulary Be? *Applied Linguistics* 11(4), 341–363, dostupné z: < <https://doi.org/10.1093/applin/11.4.341> >.

Hriadel, O. (2020): Kognitívna (ne)uzavretosť umelej inteligencie (Deep Learning), in Cintula, I. – Mydlová K. – Kalivodová, V. (eds.) *LOQUERE: interdisciplinárna doktorandská konferencia: zborník príspevkov*, Univerzita sv. Cyrila a Metoda v Trnave, 105–117.

Hsu, F-H. (2004): *Behind Deep Blue: Building the Computer that Defeated the World Chess Champion*, Princeton University Press.

Chollet, F. (2019): *Deep learning v jazyku Python*, Grada.

Kakas, A. – Michael, L. (2020): Abduction and Argumentation for Explainable Machine Learning: A Position Survey, *Cornell University arXiv.org > cs > arXiv:2010.12896*, 1–24, dostupné z: < <https://arxiv.org/abs/2010.12896> >.

Kruger, N. et al. (2013): Deep Hierarchies in the Primate Visual Cortex: What Can We Learn for Computer Vision?, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(8), 1847–1871, dostupné z: < [10.1109/TPAMI.2012.272](https://doi.org/10.1109/TPAMI.2012.272) >.

Lacker, K. (2020): Giving GPT-3 a Turing Test, *Kevin Lacker's blog* [online], 2020-07-06, [cit. 2021-03-31], dostupné z: < <https://lacker.io/ai/2020/07/06/giving-gpt-3-a-turing-test.html> >.

Levesque, H. J. (2017): *Common Sense, the Turing Test, and the Quest for Real AI*, The MIT Press.

López-Rubio, E. (2018): Computational Functionalism for the Deep Learning Era, *Minds and machines* 28, 667–688, dostupné z: < <https://doi.org/10.1007/s11023-018-9480-7> >.

Mooney R. J. (2000): Integrating Abduction and Induction in Machine Learning, in Flach, P. A. – Kakas, A. C. (eds.) *Abduction and Induction. Applied Logic Series* 18, 181–191, dostupné z: < https://doi.org/10.1007/978-94-017-0606-3_12 >.

- Moor, J. H. (2001): The Status and Future of the Turing Test, *Minds and Machines* 11, 77–93, dostupné z: < <https://doi.org/10.1023/A:1011218925467> >.
- Nguifo, E. M. (1998): Abduction and Induction in Learning Task: which needs the other? *ACADEMIA – Accelerating the world's research* 1–3, dostupné z: < https://scholar.google.com/scholar?hl=sk&as_sdt=0%2C5&q=Abduction+and+Induction+in+Learning+Task%3A+which+needs+the+other+%3F&btnG= >.
- Otterlo, M. V. (2013): A Machine Learning View on Profiling, *Privacy, Due Process and the Computational Turn The Philosophy of Law Meets the Philosophy of Technology*, 41–65, dostupné z: < <https://doi.org/10.4324/9780203427644> >.
- Parnas, D. L. (2017): The real risks of artificial intelligence, *Communications of the ACM* 60(10), 27–31, dostupné z: < <https://doi.org/10.1145/3132724> >.
- Searle, J. (1984): *Mysl, mozek a věda*, Mladá fronta, 1994.
- Schubbach, A. (2019): Judging machines: philosophical aspects of deep learning, *Synthese* 198, 1807–1827, dostupné z: < <https://doi.org/10.1007/s11229-019-02167-z> >.
- Silver, D. – Huang, A. – Maddison, C. et al. (2016): Mastering the game of Go with deep neural networks and tree search, *Nature* 529, 484–489, dostupné z: < <https://doi.org/10.1038/nature16961> >.
- Slonim, N. – Bilu, Y. – Alzate, C. et al. (2021): An autonomous debating system, *Nature* 591, 379–384, dostupné z: < <https://doi.org/10.1038/s41586-021-03215-w> >.
- Tan C. et al. (2018): A Survey on Deep Transfer Learning, in Kůrková V. et al. (eds.) *Artificial Neural Networks and Machine Learning – ICANN 2018. Lecture Notes in Computer Science* 11141, 270–279, dostupné z: < https://doi.org/10.1007/978-3-030-01424-7_27 >.
- Toledo, A. et al. (2019): Automatic Argument Quality Assessment – New Datasets and Methods, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, D19-1564, 5625–5635, dostupné z: < <https://doi.org/10.18653/v1/D19-1564> >.
- Turing, A. (1950): *Počítacie stroje a inteligencia*, Medzinárodná účastnícka spoločnosť Bradlo, 1992.
- Turing, A. (1950): Computing Machinery and Intelligence. *Mind* 59(236), 433–460, dostupné z: < <https://doi.org/10.1093/mind/LIX.236.433> >.
- Vinyals, O. et al. (2016): Matching Networks for One Shot Learning, *Advances in Neural Information Processing Systems* 29, 1–12, dostupné z: < <https://arxiv.org/abs/1606.04080> >.
- Yamins, D. – DiCarlo, J. (2016): Using goal-driven deep learning models to understand sensory cortex, *Nat Neurosci* 19, 356–365, dostupné z: < <https://doi.org/10.1038/nn.4244> >.

Zhou, Z-H. (2019): Abductive learning: towards bridging machine learning and logical reasoning, *Sci. China Inf. Sci.* 62(76101), dostupné z: < <https://doi.org/10.1007/s11432-018-9801-4> >.



Toto dílo lze užit v souladu s licenčními podmínkami Creative Commons BY-NC-ND 4.0 International (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>). Uvedené se nevztahuje na díla či prvky (např. obrazovou či fotografickou dokumentaci), které jsou v díle užity na základě smluvní licence nebo výjimky či omezení příslušných práv.
