

Môžu stroje myslieť?

Turingova imitačná hra a Searlova čínska izba sú dva z najvýznamnejších myšlienkových experimentov zaoberajúcich sa problémom „myslenia digitálnych počítačov“. Vo svojej práci sa pokúsim dokázať nasledovné: 1. Turingova imitačná hra je principiálne nezmyselná, 2. počítače nemôžu myslieť.

Základom práce je definovanie pojmu myslenia. Táto definícia nemusí byť vyčerpávajúca, stačí ak vystihuje hlavné črty toho, čo robí človek vtedy, keď hovoríme, že myslí. Pretože sama imitačná hra je založená na substitúcii digitálneho počítača (ďalej len počítača) za „dospelého“ človeka, nemusíme sa v tejto definícii zaoberať otázkou myslenia zvierat a detí, ktoré nepoznajú jazykový systém používaný v imitačnej hre, z čoho vyplýva, že nedokážu odpovedať na položené otázky.

Pracovná definícia myslenia, teda základné atribúty myslenia:

- syntax – formálne narábanie so znakmi
- sémantika – význam znakov
- interpretácia – vysvetlenie znaku z hľadiska dôležitosti, súhlasu, metaforickosti, etc.

Tejto koncepcii zhruba zodpovedá aj definícia interpretácie, ktorá má z pohľadu semiotiky tri úrovne: syntaktickú (rozoznanie znaku vzhľadom na iné znaky), sémantickú (porozumenie významu znaku) a pragmatickú.

Imitačná hra

Imitačná hra je myšlienkový experiment Alana M. Turinga, ktorý sa pýta, či je možné predstaviť si počítač, ktorý by v tejto hre obstál, čo by znamenalo, že istým spôsobom myslí. Turing svoj experiment publikuje v roku 1951, teda v čase, keď sa počítačová technika ešte len začala rozvíjať. Samozrejme, že už vtedy si mohol p r e d s t a v i ť počítač, ktorý by obstál v imitačnej hre. Táto predstava predsa nie je logicky sporná. Je však úspech počítača v imitačnej hre skutočne dôkazom jeho myslenia?

Podstata imitačnej hry: „Hrajú ju traja hráči, muž (A), žena (B) a moderátor (C), na ktorého pohlaví nezáleží. Moderátor je v inej miestnosti ako ostatné dve osoby. Cieľom hry moderátora je určiť, ktorá z dvoch osôb je muž a ktorá žena. Pozná ich ako hráča X a hráča Y a na konci hry povie „X je A a Y je B“ alebo „X je B a Y je A“. Moderátor môže hráčovi A i hráčovi B kľásť otázky. ...Cieľom hry hráča A je, aby sa hráč C pri identifikácii zmýlil. ...Cieľom hry hráča B je pomáhať moderátorovi.“ ([1], 18 - 19)

Neskôr hráča A nahradíme počítačom a budeme zisťovať, či sa moderátor bude pri rozhodovaní myliť rovnako často ako predtým. Odpovede na otázky môžu byť pravdivé aj nepravdivé.

Zmysluplnosť imitačnej hry: Úspech počítača nezávisí iba od jeho schopnosti odpovedať tak, ako by odpovedal človek. Jeho úspech je vlastne daný neúspechom moderátora, ktorý je výsledkom jeho psychologických daností a intelektuálnych schopností (či skôr neschopností). Pre vylúčenie istých typov otázok (Koľko je 16^3 ?) predpokladám, že moderátor nevie o nahradení hráča A počítačom, čo dáva počítaču vyššie šance na úspech. Moje argumenty však budú mať rovnakú silu aj bez tohto predpokladu. Je dôležité, či moderátor vie, že úlohou jedného hráča je pomáhať mu, teda hovoriť pravdu.

1. moderátor vie o hráčovi, ktorý hovorí pravdu:

a) ak počítač zvolí stratégiu, že bude dávať pravdivé odpovede, moderátor môže po dvoch otázkach určiť, kto je X a Y. Potenciálna úspešnosť počítača – žiadna.

b) ak počítač zvolí stratégiu, že bude dávať iba nepravdivé alebo pravdivé aj nepravdivé odpovede, schopnosť moderátora určiť, kto je X a Y klesne. Potenciálna úspešnosť počítača – nízka (moderátor vie, že jedna odpoveď je pravdivá).

2. moderátor nevie o hráčovi, ktorý hovorí pravdu:

- táto možnosť pri ktorejkoľvek stratégii počítača výrazne znižuje moderátorovu schopnosť s istotou určiť, kto je X a Y. Potenciálna úspešnosť počítača – vysoká. Je však závislá od osobnosti moderátora, od jeho

inteligencie, skúseností, psychologických vlastností, etc. Teda v konečnom dôsledku je výsledkom nesprávneho tipovania, teda náhody.

Turing vo svojom myšlienkovom experimente idealizoval jeho podmienky tak, že nezohľadnil subjektivitu moderátora. A práve subjektivita moderátora hrá významnú úlohu v tomto experimente, pretože ona výrazne ovplyvňuje jeho schopnosť v neevidentných prípadoch určiť, kto je X a Y. Úspech počítača v hre je mnohokrát náhodný, z čoho vyplýva, že hodnota výsledkov imitačnej hry je skreslená. Teda ak by bolo argumentom pre potvrdenie myslenia počítača určité percento úspešnosti v imitačných hrách, tento argument by bol neprijateľný. Môžete však namietnuť, že úspech počítača nie je v tom, že vyhrá imitačnú hru, ale v tom, že ju dokáže úspešne hrať bez ohľadu na záverečnú ľubovôľu moderátora pri určovaní, kto je X a Y. Tejto námietke, a teda spôsobu hry počítača, sa budem venovať v nasledujúcej časti práce.

Searlova čínska izba

Jednou z reakcií na presvedčenie, že počítače môžu myslieť je myšlienkový experiment Johna R. Searla. Jeho námietky sú nezávislé na stave techniky, pretože miera na samotnú definíciu počítača, teda na to, čo počítač je a ako ho v súčasnosti chápeme.

Podstata čínskej izby: Predstavme si program, ktorý na počítači dokáže simulovať porozumenie čínskemu jazyku. Počítač odpovedá na otázky tak, že ich porovná so svojou pamäťou a dá adekvátnu čínsku odpoveď. Rozumie však počítač čínštine tak, ako rodení Číňania? „A keď si predstavte, že by vás niekto zavrel do miestnosti, v níž je niekoľko košů naplněných čínskými znaky. A představte si, že (stejně jako já) nerozumíte čínsky, že však máte k dispozici manuál pro manipulaci s čínskými znaky, napsaný ve vašem jazyce. Pravidla přitom určují způsob manipulace se znaky čistě formálně – pouze s ohledem na jejich syntax a nikoli na jejich sémantiku. Jedno takové pravidlo např. zní: ‘Vezmi znaky škrky-škrk z koše č. 1 a polož ho vedle znaku čmrky-čmrk z koše dva 2.’ ... Předpokládejme, že znaky podávané

dovnitř se bez vašeho vědomí chápou jako 'otázky' a symboly podávané ven jako 'odpovědi'. ... A tak jste zavřeni v místnosti, skládáte znaky a podáváte je ven jako odpověď na znaky zvenčí. Avšak na základě situace, kterou jsem právě popsal, pouze manipulací formálními znaky byste se nikdy nemohli naučit čínsky.“ ([2], 33 – 34)

Rád by som sa teraz vrátil späť k Turingovej imitačnej hre. Ak prijmeme predpoklad, že počítač hrajúci túto hru je analogický čínskej izbe, vyvrátením tézy „čínska izba myslí“, vyvrátime aj tézu „počítače myslia“.

1. lingvistický argument – Searle hovorí, že operácie počítača so znakmi sú čisto formálneho charakteru. Postupnosť týchto operácií je vykonávaná pomocou abstraktných symbolov, je iba radom núl a jednotiek. „Symboly ovšem nemají ani význam, ani sémantický obsah – na nic se nevztahují.“ ([2], 32) Zo semiotického hľadiska binárny kód, v ktorom pracujú počítače je rovnako ako jazyk, v ktorom sa dorozumievajú ľudia, znakový systém s úrovňou druhej artikulácie, ibaže binárny kód má na rozdiel od nášho jazyka iba túto úroveň artikulácie. Správa je zo semiotického hľadiska artikulovaná vtedy, keď môže byť rozložená na elementy a každý z týchto elementov je signifikantný. Na úrovni prvej artikulácie sa systém skladá z najmenších dostupných zmysluplných jednotiek, ktoré sú úplnými znakmi, teda pozostávajú z označovaného a označujúceho. Na úrovni druhej artikulácie je systém rozložiteľný na minimálne funkčné jednotky, ktorým chýba význam. Tieto čisto odlišujúce štrukturálne jednotky sú opakujúcimi sa znakmi v kóde. Systém musí mať prvú úroveň artikulácie, aby tieto nižšie jednotky mohli byť kombinované do zmysluplných znakov. Binárny kód má iba druhú úroveň artikulácie, a preto počítače, ktoré pracujú na jeho báze, nemajú sémantiku. Keďže jedným z atribútov myslenia v našej pracovnej definícii bola sémantika, vyplýva z toho, že počítače nemôžu myslieť. Okrem toho absencia sémantiky úplne vylučuje schopnosť interpretácie znakov, čo bol náš tretí atribút. „Mysl má víc než jen syntax, má i sémantiku. A žádný počítačový program nikdy nebude mít mysl prostě proto, že je pouze syntaktický, zatímco mysl není jen syntaktická.“ ([2], 33) Osoba v čínskej izbe má rovnako ako počítač formálne pravidlá pre operácie so znakmi, ale

nikdy z toho nepochopí význam týchto znakov, teda nebude mať sémantiku, ale iba syntax. Tento argument sa nevzťahuje primárne iba na počítače hrajúce Turingovu imitačnú hru, ale na všetky digitálne počítače. Spôsob, akým môže hrať počítač imitačnú hru, popíšem v druhom argumente.

2. argument simulácie – podstatou tohto argumentu je, že počítače nemyslia, ale iba simulujú (napodobňujú) myslenie. Rozdiel medzi simuláciou a skutočnosťou môže byť z funkcionálneho hľadiska nepodstatný, pretože počítač môže niekedy nasimulovať situáciu alebo výsledok procesu tak, ako sa neskôr stane v skutočnosti, ale z ontologického a epistemologického hľadiska je tento rozdiel fundamentálny. Z tohto pohľadu sa môže počítač stať výborným iluzionistom myslenia, ale nikdy nebude myslieť. To je dôvod, prečo môže v celku jednoduchým spôsobom uspieť v imitačnej hre.

Jazyk má dve dimenzie (osi) – syntagmatickú a paradigmatickú. Syntagmatická os je založená na kombinácii znakov. Ide v podstate o syntax – spájanie slov tak, aby vytvorili gramaticky správnu vetu. Paradigmatická os je naopak založená na výbere znakov, ktoré budú dosadené na určité miesto v syntagme. V obyčajnom subjekt–predikátovom výroku je syntagmatická os napr. podstatné meno (subjekt) + kopula (byť) + prídavné meno (predikát) a paradigmatická množina znakov z definovanej kategórie, v ktorej každý označuje niečo iné, napr. prídavné mená (malý, veľký, žltý). Syntagmatická os je v skutočnosti problémom formálneho kalkulu, a preto nie je pre počítač nemožné, aby na otázky v imitačnej hre odpovedal gramaticky vhodnými odpoveďami. Dokáže však počítač vyberať správne slová z paradigmatickej osi? Výber slov z paradigmatickej osi je závislý od kontextu, ktorého pochopenie si vyžaduje schopnosť interpretácie. Ako som však ukázal v lingvistickom argumente, počítače touto schopnosťou nedisponujú. Ako je potom možné, aby počítač hral imitačnú hru? Počítač v podstate odpovedá iba na dva druhy otázok – zisťovacie a doplňovacie. Jeho prvou úlohou pri odpovedi v imitačnej hre je teda rozpoznanie druhu otázky, čo pri dobrom naprogramovaní nemusí pôsobiť ťažkosti. Vhodnou odpoveďou na zisťovaciu otázku je *áno* alebo *nie*, teda *je pravda, že ...* alebo *nie je pravda, že* Keďže počítač môže v imitačnej hre klamať, je odpoveď na zisťovaciu otázku

problémom jednoduchého formálneho kalkulu – $a \rightarrow b$ v c (Ak zisťovacia otázka, tak áno alebo nie), pričom na pravdivostnej hodnote odpovede nezáleží. V prípade zisťovacej otázky sa počítač pohybuje iba na úrovni syntagmatickej osi. V prípade doplňovacej otázky môžu nastať dva prípady. Počítač má odpoveď na otázku uloženú v pamäti alebo ju môže vygenerovať algoritmom. Teda počítač pracuje ako encyklopédia alebo kalkulačka. Tento proces je tiež iba problémom formálneho priradenia. V poslednom prípade, keď počítač nemá v pamäti uloženú odpoveď na doplňovaciu otázku, musí vyberať slová z paradigmatickej osi. Ale keďže počítač v imitačnej hre môže odpovedať aj nepravdivo, stačí, ak vyberie hociktoré slovo z vhodnej paradigmatickej množiny. Kritériom výberu vhodnej paradigmatickej množiny je syntagmatické postavenie slova a výber konkrétneho slova je u ž potom náhodný. Celý tento proces je dobre formalizovateľný a schopnosť interpretácie znakov v ňom nehrá žiadnu úlohu. Týmto spôsobom môže počítač v imitačnej hre dobre simulovať ľudské myslenie.

Lingvistický argument a argument simulácie stačia na to, aby sme mohli tvrdiť, že neplatnosť tézy „čínska izba myslí“ je dokázaná, z čoho vyplýva aj neplatnosť tézy „počítače myslia“.

Literatúra

1. Turing, A.: Počítacie stroje a inteligencia. In: Myseľ, telo, stroj. Bradlo 1992
2. Searle, J. R.: Mysl, mozek a věda. Praha, Mladá Fronta 1994