

Turingov vplyv na súčasné učiace sa systémy

Ondrej Hriadel

Ondrej Hriadel: Turingov vplyv na súčasné učiace sa systémy [Turing's influence on current learning systems] In: *Ostium*, vol. 19, 2023, no. 3.



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Turing's influence on current learning systems

Alan Turing did not just design the Turing test but had other ideas about machine intelligence. He realizes that the exact computation that applies to the Turing machine won't be sufficient for artificial intelligence (AI). The connectionist approach to AI also supports this idea in which he analyzes the attributes that machines should have to be intelligent. Turing also contemplates the concept of machine learning because the ability to learn could help machines to work freely without being reliant on computation. This approach would allow us to pinpoint the signs of intelligent behavior. Regarding the contemporary AI and deep learning based on learning algorithms, we can assume that Turing's ideas about connectionism and learning are relevant and impact current learning systems. This point of view, despite its obvious technical implications, is also important philosophically because it represents the opposite of classic (symbolic) AI. Connectionist AI could be more similar to machine intelligence than to the classical.

Keywords: Turing, learning, learning machines, deep learning (DL), Turing machine, connectionism

Úvod[1]

Popri Turingovom teste je Alan Turing uznávaný ako otec počítačov, keďže sa v prvej polovici minulého storočia významne podieľal na vzniku prvých výpočtových strojov z technického i teoretického hľadiska. Za najväčšie Turingove úspechy v počítačovej oblasti je možné považovať vytvorenie stroja na dešifrovanie nemeckej Enigmy a ACE[2], ktoré dopomohli vývoju výpočtovej techniky a ukončeniu vojny. Okrem toho sú dôležité i jeho teoretické poznatky prezentované v prácach *Computing Machinery and Intelligence*[3] a *Intelligent machinery*[4], keďže poukázali na potenciálnu podobu novodobých počítačov, ale aj umelej inteligencie (ďalej len UI). Keďže sa Turingove myšlienky nezameriavali iba na konštrukciu počítačov, ale aj na systémy, ktoré môžu disponovať ľudskými vlastnosťami, výskum nezostal iba v oblasti technických vied, ale rozšíril sa do humanitných odborov vďaka myšlienkam o inteligentných strojoch.

V súčasnosti sa názory na Turingov vplyv vo vývoji dnešných počítačov do určitej miery rozchádzajú. Z technického hľadiska sa môže zdať, že sa vytráca záujem o túto problematiku. Príznakom toho môže byť, že Turingovo definovanie *Turingovho stroja* a *Univerzálneho Turingovho stroja* ostalo iba na teoretickej úrovni. Ako podotýkajú aj Copeland a Proudfootová: „Často sa hovorí, že okrem špecifikácie Univerzálneho Turingovho stroja, Turing zohral malú alebo žiadnu úlohu vo

vývoji počítačov.“[5] I keď by bolo možné tento názor považovať za pravdivý, konštatuje iba technickú úroveň počítačov, bez ohľadu na vývoj UI aj v iných oblastiach. Turingove nápady boli na jednej strane síce teoretické, na druhej strane mali interdisciplinárny presah. Abramson uvádza, že Turing je pre kognitívnych vedcov významný z dvoch dôvodov. „Po prvé, je jednou z kľúčových postáv pri vynájdení moderného pojmu univerzálneho počítača, ktorý je kľúčový pre kognitívnu revolúciu. Po druhé, Turingove názory na možnosť strojovej inteligencie boli inšpirujúce pre generácie odborníkov v oblasti UI.“[6] Abramson primárne spomína kognitívne vedy, avšak využíva pojem strojová inteligencia, ktorá je dôležitým problémom i vo filozofii. Otázky možnosti inteligentných strojov sú rozoberané už od vzniku klasickej UI. V posledných rokoch sa však problematika môže posunúť ďalej vďaka vývoju v strojovom učení (ďalej len ML) a hlbokom učení (ďalej len DL)[7], ktoré sú v súčasnosti jednou z najdôležitejších oblastí UI. Zároveň prichádzajú tieto technológie s kognitívnymi atribútmi, ktoré sú podobné človeku, hlavne v prípade učenia.

V súvislosti s Turingom sú jeho myšlienky o strojovej inteligencii často spájané iba s Turingovým testom, dôležitým prínosom sú však aj myšlienky o učiacich sa strojoch. Pravdepodobne si uvedomoval, že v budúcnosti nebude postačovať iba samotný výkon počítačov na to, aby sa mohli stať inteligentnými, pretože tvrdí: „Inteligentné správanie spočíva v odklone od úplne disciplinovaného správania založeného na výpočtoch. Avšak ide o odklon pomerne mierny, ktorý nevedie k náhodnému správaniu alebo k nezmyselným opakujúcim sa cyklom.“[8] Hoci vyzerá Turingovo vyjadrenie na prvý pohľad vágne a neiste, v jeho prácach možno nájsť štruktúrované nápady o učiacich sa strojoch, ktoré sa úplne kontrolovanému správaniu vymykajú. Vzhľadom na túto skutočnosť je cieľom tejto práce poukázať na to, že Turingove predpoklady môžu mať vplyv aj na súčasné systémy UI. Konkrétne sa budem zameriavať na Turingov stroj, Turingov pohľad na konekcionizmus s neorganizovanými strojmi a predpoklady o učení v spojitosti s DL.

1. Turingov stroj

Turingov stroj disponujúci nekonečnou pamäťou (páskou)[9] je teoretickým konceptom najvýkonnejšieho počítača, pretože by mu jeho pamäť umožňovala z počítačného hľadiska fungovať neustále, až pokiaľ by nevyriešil problém. Dokonca už súčasné počítače sú výpočtovo rýchlejšie ako ľudský mozog (napr. superpočítač Summit[10]) a mnohé systémy spĺňajú Turingovu úplnosť.[11] To by mohlo predznamenať, že Turingov prístup bol správny vo vývoji počítačov, ktoré by boli vhodné i pre pokrok v oblasti UI. Napriek rýchlosti podobných zariadení, ako aj Turingovho stroja sú v súvislosti s DL prítomné limitácie, ktoré nekorešponujú s možnosťami učiacich sa systémov. Primárnym problémom je determinovanosť programového nastavenia a ako uvádzajú Hernández-Orallo - Dowe: „Turingov stroj nie je interaktívny. Kognitívne schopnosti sa však zvyčajne spájajú s interaktívnymi systémami.“[12] Vzhľadom na to, že DL je významné kognitívnymi možnosťami dôležitými pri procese tréningu (učenie), vyvstáva otázka, či má Turingov stroj nejaký vplyv na súčasnú podobu týchto systémov. Prípadne či ide o teoretický návrh, ktorý vyniká iba potenciálne nekonečnou pamäťou. Konkrétnejšie sa na tento problém možno pozrieť prostredníctvom systémov DL, ktoré sa nechali inšpirovať Turingovým konceptom.

1.1 Modely inšpirované Turingovým strojom

Za najbližšie projekty Turingovmu stroju je možné považovať *Super-Turingov model*[13] Havy Siegelmannovej a *Turingov neurónový stroj*[14] Alex Gravesa a kol. V oboch prípadoch ide síce o technické práce, nachádzajú sa však v nich implikácie, na základe ktorých je možné analyzovať relevanciu pôvodného návrhu Turingovho stroja.

Siegelmannová vo svojom modeli taktiež pracuje s možnosťou počítačne neobmedzeného systému.[15] Napriek matematickému predpokladu narazila na problém, ktorý sa striktne netýka počítačnej sily Turingovho stroja, ale naopak limitácie, že okrem výpočtového výkonu nedisponuje takými atribútmi, akými DL disponuje. Konkrétne hovorí o učení a adaptácii, ktoré sú v Super-Turingovom modeli aplikované, aby systém nebol determinovaný iba na úlohy, na ktoré bol vopred

naprogramovaný. To znamená, že rieši kľúčový problém Turingovho stroja, teda že vykonáva úlohy bez prispôsobovania, iba tak, ako mal presne určené. Na margo svojho modelu Siegelmannová dokonca tvrdí, že „Super-Turing je úplne odlišný od Turingovho výpočtu. Hierarchia výpočtových schopností priamo súvisí so schopnosťou prispôbiť sa okolitému prostrediu s ekvivalentnou bohatosťou informácií, ktoré sú k dispozícii pre výpočtové operácie.“[\[16\]](#) Pri porovnaní oboch modelov je zrejmé, že vlastnosti Siegelmannovej modelu prevyšujú možnosti, ktoré ponúka iba výpočtový výkon. Zároveň na základe jej vyjadrenia je možné usúdiť, že Super-Turing model rieši aj Hernández-Orallovo – Doweovo podnet o nemožnosti interakcie, keďže sa dokáže prispôbiť okolitému prostrediu (riešenej úlohe).

Turingov neurónový stroj na rozdiel od Super-Turingového modelu okrem uvádzaných vlastností prináša dodatočný typ pamäte integrovaný do neurónovej siete, ktorý má podobnosť s ľudskou pracovnou pamäťou. Už predchádzajúci model poukázal na možnú nedostatočnosť Turingovho stroja ako koncept pre DL systémy. Turingov neurónový stroj sa však v tomto smere posúva ešte o krok ďalej vďaka inšpirácii a do určitej miery technickému prevedeniu biologickej pamäte. Napriek tomu autori nemajú v porovnaní so Siegelmannovou negatívny prístup k využitiu a inšpirácii digitálnych počítačov, ale považujú svoje zariadenie skôr za kombináciu oboch. „Neurónový Turingov stroj je architektúra neurónovej siete, ktorá čerpá inšpiráciu z modelov biologickej pracovnej pamäte a z konštrukcie digitálnych počítačov.“[\[17\]](#) I keď je potrebné dodať, že pamäť v ich modeli je v porovnaní s digitálnym počítačom prispôbivejšia.[\[18\]](#)

Na základe oboch prípadov možno usúdiť, že sa návrh Turingovho stroja zdá byť nedostatočný pre pokrok v UI. Za predpokladu, že by bol narastajúci výpočtový výkon jediným kľúčovým faktorom, mnohé súčasné počítače by túto podmienku spĺňali. Výkon však nemusí byť najzásadnejším problémom, keďže mnohé systémy, ako aj tie, čo boli uvedené, spĺňajú Turingovu úplnosť. Inšpirácia Turingovým strojom z toho dôvodu súvisí iba s tým, že sú výpočtovo dostatočné, aby boli schopné vykonávať ďalšie funkcie ako učiace sa systémy. Schopnosti ako učenie, adaptácia, interakcia v trénoch úlohách alebo dokonca prítomnosť počítačovej verzie pracovnej pamäte síce nemusia automaticky predpokladať inteligenciu v DL (najmä v porovnaní s človekom). Napriek tomu tieto atribúty predstavujú iný smer ako pristupovať k problému strojovej inteligencie, ktorý nezávisí iba od komputácie.

Kugel, ktorý ešte pred rozšírením DL súhlasil, že komputácia pravdepodobne nebude dostatočná na dosiahnutie strojovej inteligencie, aj tak Turinga obhajuje. Konkrétne sa zameriava na to, že si Turing uvedomoval nedostatočnosť komputačného prístupu: „Myslím, že chcel povedať. Nebude možné naprogramovať stroje obmedzené na výpočty tak, aby boli skutočne inteligentné.“[\[19\]](#) S Kugelom je možné súhlasiť v tom, že Turingova predstava inteligentných strojov nespočívala iba v návrhoch, akým je aj Turingov stroj. Preto je dôležité na ďalšie Turingove myšlienky o konekcionizme a učení nahliadať bez ohľadu na jeho koncept výpočtového stroja.

2. Konekcionizmus a neorganizované stroje

Hoci Turing nebol prvý, kto pracoval s konceptom neurónových sietí[\[20\]](#), jeho práca *Intelligent machinery* prináša návrhy, ako efektívne zapojiť (umelé) neuróny do štruktúry, ktorá môže byť učiacim sa systémom. Copeland – Proudfootová tvrdia, že „Turing bol pravdepodobne prvým človekom, ktorý uvažoval o vytvorení výpočtových strojov z jednoduchých prvkov podobných neurónom, ktoré by boli do značnej miery náhodne prepojené do siete“.[\[21\]](#) Z technického hľadiska možno tieto návrhy považovať za predchodcov súčasných neurónových sietí, a teda i určitého smeru UI. To však nemusí byť jediný význam, ktorý jeho koncepty priniesli. Z filozofického hľadiska by ich bolo možné považovať za protipól klasickej symbolickej UI, založenej na konekcionizme prinášajúci iný prístup k strojovej inteligencii v porovnaní s jej náprotivkom.

Turingovou predstavou bolo vytvoriť stroje, ktoré by boli konštruované náhodným spôsobom

prostredníctvom neurónových spojení namiesto toho, aby boli presne naprogramované. Takéto stroje by podľa neho neboli determinované, naopak, disponovali by náhodným usporiadaním, boli by neorganizované. Medzi mnohými návrhmi je pravdepodobnejšie najdôležitejší tzv. neorganizovaný stroj typu B kvôli modifikovateľnosti, ktorý má spoločné prvky so súčasným DL, ako aj otázkou strojovej inteligencie. Vzhľadom na tento problém a jeho prepojenie s konekciónizmom je však potrebné začať Turingovou inšpiráciou mozgom.

I keď Turing uvažoval o možnosti nasimulovať celého človeka, aby sa dal vytvoriť inteligentný stroj, rýchlo ju zavrhol. „Namiesto toho sa pokúsme zistiť, čo sa dá urobiť s ‚mozgom‘, ktorý je viac-menej bez tela a poskytuje nanajvýš orgány zraku, reči a sluchu. Potom stojíme pred problémom, ako nájsť vhodné odvetvia, aby stroj mohol uplatniť svoje schopnosti.“^[22] Človeku síce jeho mozog umožňuje vykonávať rôzne intelektuálne úkony (rozhodovať, riešiť problémy atď.), avšak aj ovládať a zabezpečovať fungovanie tela. Redukcia na neurónové zapojenie, ktoré by slúžilo iba na riešenie abstraktných činností bez potreby kontrolovať iné funkcie, sa zdá byť správnu cestou, a to najmä v prípade neurónových sietí. Inými slovami, mnohé učiace sa systémy fungujú iba na úrovni počítačových algoritmov bez integrácie do ďalších zariadení.^[23] Napriek tomu sú schopné vykonávať úlohy ako hranie hier, práca s jazykom či kryptografia, o ktorých hovoril aj Turing. Vzhľadom na takýto konekciónistický prístup, a teda možnosť aplikovať neurónové siete DL na rôzne úlohy je otázne, či možno hovoriť o nich ako o inteligentných strojoch. Teucher upozorňuje a taktiež sa pýta: „Niektorí významní kritici dokonca tvrdili, že skutočne inteligentný stroj sa nedá vytvoriť. Čo je však inteligencia a čo môžu inteligentné stroje robiť?“^[24] Pesimistický názor o nemožnosti vytvoriť inteligentný stroj možno považovať za relevantný, vzhľadom na to, že i súčasné DL je založené na umelých neurónových sieťach. Na druhej strane, schopnosti DL prevyšujú determinovanú klasickú UI a hoci v súvislosti s Turingom nie je možné hovoriť o konkrétnej definícii alebo vysvetlení strojovej inteligencie, pripadá do úvahy, že naznačil cestu, ako sa takému stavu priblížiť prostredníctvom modifikovateľnosti a organizovaní neorganizovaných strojov.

2.1 Modifikovateľnosť a organizovateľnosť

Prostredníctvom oboch aspektov sa Turing nesnažil preukázať iba na možnosť inteligentných strojov, ale aj na ich univerzálnosť. I keď môže byť prepojenie týchto aspektov zrejmé a i v súvislosti s UI žiadané, v skutočnosti je tento stav pravdepodobne ďaleko. Avšak i za takýchto podmienok je možné v jeho predpokladoch nájsť implikácie, ktoré by mohli predznamenávať pokrok strojovej inteligencie.

Prvý spôsob, modifikovateľnosť, Turing spája so *samo-modifikovateľnosťou* a *modifikovateľnosťou prostredníctvom interferencie*.^[25] Hoci opísal tieto koncepty na neexistujúcom stroji, v súčasnosti sú aplikovateľné. Rozmanitosť dnešných neurónových sietí a DL algoritmov je síce značná, základné metódy ako učenie s učiteľom (ďalej len SL), učenie bez učiteľa (ďalej len UL) a učenie s posilňovaním (ďalej len RL) však postačujú na preukázanie, kam asi smerovali Turingove predpoklady o konekciónistickej UI.

Prvé dve metódy sú zrejším príkladom samo-modifikovateľnosti (UL) a modifikovateľnosti prostredníctvom interferencie (SL). Obe metódy sú síce závislé od značného množstva trébovaných dát, ich rozdiely sú však v súvislosti s Turingom kľúčovejšie. Keďže SL závisí od presne definovaných dát (príkladov, na ktorých sa má trébovať), jeho interferenciou sú samotné dáta, od ktorých závisí výsledok. Môže sa síce zdať, že takýto prístup k modifikovateľnosti v DL má vlastne blízko ku klasickej UI^[26], avšak je veľmi podobná tomu, čo Turing nazval *papierovou interferenciou* „spočívajúcou v samotnom odovzdávaní informácií do stroja, ktoré menia jeho správanie.“^[27] Dokonca možno dodať, že by bez takejto interferencie nemohol vzniknúť správny výsledok.^[28] UL síce taktiež potrebuje na trébovanie množstvo dát, metóda však slúži na nájdenie štruktúry či anomálie v určitom probléme, namiesto trébovania konkrétnych vzorových dát. I keď ide o určitú interferenciu vo vstupných dátach, samorganizácia^[29], ktorou UL disponuje, zodpovedá Turingovmu pomenovaniu samo-modifikovateľnosti, pretože si určuje, akým spôsobom problém vyrieši. Metódu

RL by Turing možno považoval za najviac samo-modifikovateľný neorganizovaný stroj, pretože zmeny v procese tréovania závisia od virtuálneho konatela interagujúceho s prostredím riešeného problému a bez nutnosti ďalšej interferencie (s výnimkou základných dát tréovanej úlohy). Zároveň ide o postup, ktorého základy (princíp odmien a trestov) považoval za spôsob, ako skonštruovať učiaci sa stroj, pričom tento princíp funguje aj v dnešnom RL.[\[30\]](#)

Druhý spôsob, organizovanie neorganizovaných strojov, je u Turinga taktiež spojený s učením, bude mu však venovaná samostatná časť. Z konekcionistického hľadiska je dôležité, že považoval za neorganizovaný stroj mozog dieťaťa, ktorý by mal byť súčasne s tým aj univerzálnym strojom. Na problém možno nahliadať z dvoch strán - vhodnosť detského mozgu ako návrh neorganizovaného stroja a na samostatnú univerzálnosť.

S myšlienkou použiť detský mozog ako návrh pre neorganizované stroje prišiel Turing preto, lebo považoval mozog dospelého človeka za zťažený vplyvom rôznorodého výcviku (uvádza napr. učenie sa jazykov). Na základe toho usúdil, že detský mozog týmto vplyvom nebol ešte vystavený a označil ho za (zatiaľ) nekoordinovaný. Dieťa s počiatočným stavom mozgu ešte len začína prijímať a spracovávať nové informácie, ktoré sú spočiatku neorganizované. Problémom je, či takýto predobraz mozgovej neorganizovanej siete postačuje ako základ pre tie umelé. Teuscher považuje takýto prístup za pomerne naivný. „Toto hľadisko by sa určite dalo spochybníť, pretože do značnej miery zjednodušuje mozog. I keď z biologického hľadiska existuje dobrý dôvod považovať rastúcu kôru za takmer neorganizovanú sieť.“[\[31\]](#) Teuscher svojím vyjadrením apeluje na zjednodušené neurónové zapojenia v umelých sieťach v porovnaní s biologickými. S tým je potrebné súhlasiť s ohľadom na akýkoľvek typ siete.[\[32\]](#) Na druhej strane, Turing sa zameriaval iba na koncept, ako by bolo tieto stroje možné navrhnuť. V takom prípade nie je nutné takýto prístup úplne zavrhnúť. Za predpokladu, že neurónová sieť ešte nie je natréovaná, čiže nemá pri vstupe žiadne dáta, tak sa nachádza v neorganizovanom stave. Až postupným spravovaním dát a úpravou parametrov siete (váh) je generovaná presnejšia reprezentácia, a tým sa stáva sieť organizovanejšia. Hoci Turing nemal príležitosť podrobnejšie testovať umelé neurónové siete a ich súčasná forma sa stále nevyrovná úplnej simulácii mozgu, istú podobnosť v Turingovom koncepte s aktuálnymi verziami však nemožno úplne vylúčiť.

Ako už bolo naznačené, Turingovo chápanie univerzálnosti je i v súvislosti s DL problematické, pretože mal síce jednoduchú, pritom však ťažko dosiahnuteľnú predstavu. „Význam univerzálného stroja je jasný. Nepotrebuje mať množstvo rôznych strojov, ktoré vykonávajú rôzne úlohy. Postačí jeden.“[\[33\]](#) Turing v tomto bode síce nehovoril o všeobecnej inteligencii. Minimálne však chcel, aby sa k tomu neorganizované stroje požadovanou univerzálnosťou priblížili. Ani o súčasnom DL nie je možné hovoriť ako o úplne univerzálnom, pretože sa radí do úzkej UI. Inak povedané, DL je v reálnom čase možné natréovať iba na jednu konkrétnu úlohu, ktorú je schopná vykonávať podobne ako Turingove siete typu B.[\[34\]](#) Vzhľadom na to nemožno hovoriť o úplnej univerzálnosti, je však možné predpokladať strednú cestu. DL i siete typu B sú pripravené na vykonávanie konkrétnej úlohy až po tréovaní, vtedy sa stanú uzavretými. Pred tréningom sú však variabilné, keďže závisia od vstupných dát. Druh vykonávanej úlohy môže byť tým pádom rôznorodý (spracovanie obrazu, zvuku, textu atď.[\[35\]](#)). Z toho sa dá usúdiť, že síce nejde o univerzálnosť, ako by si to predstavoval Turing. Ide však o prístup, ktorý sa k nej približuje, keďže ju možno považovať za neorganizovanú univerzálnosť.

Turingove konekcionistické návrhy o neorganizovaných strojoch ukazujú, akým smerom začal a postupoval vývoj neurónových sietí a teda i posun od klasickej UI k UI založenej na konekcionizme. I keď nemožno v súvislosti so súčasným DL so všetkým súhlasiť. V značnej miere ide o nápady, ktoré by mali mať presah i mimo technické vedy, nakoľko ukazujú základ úplne odlišnej UI. Z hľadiska strojovej inteligencie ide o prvé kroky, prostredníctvom ktorých sa učiace systémy oddaľujú od determinovaných programov. Možnosti neorganizovaných strojov, ktoré Turing predstavil v podobe

modifikovateľnosti a organizovateľnosti naznačujú v súvislosti so súčasným DL posun k inteligentnejším strojom. V prípade nápadu o organizovaní strojov síce nie je všetko v súlade s Turingovými predpokladmi, avšak návrh o učení spájanom s organizovaním môže byť dôležitým aspektom v DL.

3. Učiace sa stroje

Turing si uvedomoval, že konštrukcia inteligentných strojov bude vyžadovať viac ako komputáciu. Z tohto dôvodu hľadal spôsob, akým by sa stroje mohli priblížiť človeku minimálne v určitých kognitívnych schopnostiach. I keď bolo jeho konečným cieľom vytvoriť stroj porovnateľný s ľudskou inteligenciou, potreboval na začiatok vlastnosť – učenie –, prostredníctvom ktorej by stroje zbavil úplnej závislosti od komputácie a umožnil im tak vykonávať činnosti korešpondujúce s inteligentným správaním. Učenie teda využil ako prostriedok k návrhu strojov, ktoré by neboli závislé od presne definovaných programov, ale dotvárané ďalšími prijatými (naučenými) informáciami. V súčasnosti je ML a DL dôležitou podkategóriou UI a tréningovanie (učenie v DL) primárnym spôsobom, ako dnešné systémy dokážu riešiť kognitívne náročnejšie úlohy. V súvislosti s Turingom je teda otázne, či jeho predpoklady majú implikácie a vplyv aj na súčasné systémy využívajúce učiace sa algoritmy.

Keďže Turing nepovažoval myseľ dospelého človeka za vhodný základ pre učiaci sa stroj, navrhol dieťa, ktoré na rozdiel od dospelého ešte nemá žiadne teoretické informácie a ani nebolo vystavené vonkajšiemu svetu. „Prečo by sme sa nemohli pokúsiť vytvoriť namiesto programu simulujúceho myseľ dospelého človeka, program, ktorý by simuloval detskú myseľ? Ak by sa potom takáto simulácia podrobila vhodnému vzdelávaniu, získal by sa mozog dospelého človeka. Pravdepodobne je detský mozog niečo ako zošit, ktorý bol práve zakúpený. Málo mechanizmov a veľa prázdnych listov.“[\[36\]](#) Turingovi sa zdala detská myseľ ako ideálny nástroj s prázdnu pamäťou, do ktorej možno ľubovoľne zapisovať a upravovať ju. Pamäť síce problém predstavovať nemusí, iné mechanizmy však áno.

Groszová[\[37\]](#) uvádza, že na základe dnešných výskumov z psychológie, neurovedy a pedagogiky nemožno považovať deti za dospelých jedincov, ktorí iba potrebujú doplniť fakty a vytvoriť si skúsenosť. Musia byť aktívnymi účastníkmi učebného procesu zdieľajúcimi nové nápady spolu s interakciou s inými ľuďmi. Hall[\[38\]](#) taktiež dodáva, že aj dospelá ľudská myseľ zostáva schopná učiť sa, usudzovať a vykonávať ďalšie kognitívne aktivity. Vzhľadom na názor oboch autorov možno usúdiť, že koncept detskej mysle nemusí korešpondovať s mechanizmami, ktoré objasňuje súčasná veda. Inými slovami, iba dodávanie informácií do prázdnej schránky nemusí byť dostatočné na vytvorenie učiaceho sa stroja, ktorý by bol porovnateľný so schopnosťami (po natréningovaní) dospelého človeka. Dokonca by mohlo byť irelevantné, aká simulácia mysle by sa využila ako základ pre učiaci stroj, pokiaľ splní kognitívne kritériá s výnimkou prázdnej pamäte u detskej mysle. Za predpokladu, že sa zohľadnia súčasné učebné mechanizmy, Turingov predpoklad nepostačuje na vytvorenie učiaceho sa stroja so všetkými schopnosťami človeka. Na druhej strane sa však na problém možno pozrieť v súvislosti s DL a návrh detskej mysle prirovnať k RL konateľovi. I keď je DL v porovnaní s človekom (z hľadiska učenia) uzavretým systémom, a tým pádom nespĺňa Turingovu požiadavku na univerzálnosť, ide o prístup ako zreteľnejšie ukázať jeho vplyv na súčasne učiace sa algoritmy. Iným východiskom by bolo zamietnutie Turingovho prístupu pre (ani dnes) neuskutočniteľný cieľ o všestrannom učiacom sa stroji.

3.1 Turing a učenie v DL

Turing rozdelil problém učenia do dvoch častí – „Program dieťaťa a vzdelávací proces.“[\[39\]](#) Podľa Teuchera sa aj dnes dá toto rozdelenie definovať. „Program dieťaťa zodpovedá počiatočnej architektúre siete a počiatočným hodnotám, zatiaľ čo vzdelávací proces zodpovedá učebnému algoritmu, ktorý modifikuje („trénuje“) sieť.“[\[40\]](#) Napriek tomu, že je DL metód viacero, SL a UL sú primárne štatistické metódy, prostredníctvom ktorých sa algoritmy učia zo vzoriek dát. Na rozdiel od nich, RL disponuje konateľom interagujúcim s prostredím riešenej úlohy, čiže ide o interaktívny

prístup. Toto je dôležité v súvislosti so zložkami, ktorými by mal podľa Turinga disponovať učiaci sa stroj.

„ - Počiatkový stav mysle, napríklad pri narodení,

- vzdelanie, ktorému bola vystavená,

- iné skúsenosti, ktoré nemožno označiť za vzdelanie, ktorej bola vystavená.“[\[41\]](#)

V prvom bode je síce napísané myseľ, avšak pri porovnaní s prvou časťou učenia - detským programom - je zrejmé, že ide o iniciačný program, alebo v prípade DL učiaci sa algoritmus, ako aj naznačil Teuscher. Vzdelávanie, teda tréning, predstavujú v prípade RL iniciačné dáta riešeného problému. Požiadavka vstupných dát je nutná u hocikakého DL algoritmu. Výhodou RL je, že proces tréningu siete vyžaduje vstupné dáta, využíva ich však na interaktívne riešenie problému. To priamo korešponduje s tretím bodom. Tým, že konateľ interaguje s prostredím, vykonáva ďalšie kroky na splnenie úlohy a touto cestou si vytvára skúsenosť potrebnú na finálne riešenie úlohy alebo aj jej opätovné vykonávanie. Zároveň je nutné zmieniť, že tento proces nemožno prisúdiť druhej zložke „vzdelávaniu, ktorému bola vystavená“, keďže ho RL realizuje samostatne. Ako aj Hernández-Orallo - Dowe uvádzajú: „Z určitého objemu informácií dokáže učiaci sa konateľ získať schopnosť namiesto toho, aby mu bola vopred naprogramovaná.“[\[42\]](#) Inými slovami, učiaci sa systém dokáže samostatne nadobudnúť schopnosť, ktorú môže ďalej využívať (napr. hrať šach). Takýmto prístupom sa teda DL môže nielen učiť, ale aj interagovať a získavať skúsenosti. Na podporu toho, že RL metóda je najbližšie Turingovej predstave učiaceho sa stroja nasvedčuje aj použitie postupu odmien a trestov, ktorú Turing opísal už v minulom storočí.

Už bolo naznačené, že DL dokáže riešiť určité konkrétne činnosti, ako je šach. Turing sa taktiež na konci svojej práce zamýšľal, aké úlohy by mal potenciálny učiaci sa stroj vykonávať. „Mnoho ľudí si myslí, že najlepšia by bola veľmi abstraktná činnosť, napríklad hranie šachu.“[\[43\]](#) Samozrejme je potrebné uznať, že Turing v tomto smere usudzoval správne. RL je dokonca zameraná na riešenie podobných strategických hier. Na druhej strane je zaujímavé, že sa v tomto predpoklade odkláňa od univerzálneho stroja, keďže by v takomto prípade nezáležalo na učení konkrétnej činnosti. Turing však nad učením konkrétnych činností uvažoval už pri tvorbe ACE.[\[44\]](#) Tento prístup by mohol naznačovať, že si uvedomoval nemožnosť dosiahnuť univerzálny učiaci sa stroj. Teda i predpoklad o detskej mysli mohol korešpondovať iba s nápadom prázdnej pamäte (nepopísaný zošit), i keď by bolo samotné učenie obmedzené na jednotlivé činnosti namiesto všestranného použitia.

Predpoklady o učiacich strojoch, ktoré Turing predstavil, síce úplne nezodpovedajú súčasným učiacim sa systémom, najmä v prípade detskej mysle a jej univerzálnosti. Na druhej strane zložky, ktoré navrhoval, je možné pripodobniť DL metódam schopným učiť sa interaktívne a so skúsenosťou, i keď bez univerzálnosti, v podobe v akej si ju Turing pravdepodobne predstavoval. V prípade opätovnej otázky - čo robia inteligentné stroje - by v súvislosti s jeho prácou bolo možné odpovedať, že sa neorganizované stroje stavajú učením organizovanými, a teda i učiacimi. Táto vlastnosť im umožňuje vykonávať úlohy interaktívne a s využitím skúsenosti. Miłkowski dokonca argumentuje, že Turingove chápanie inteligencie nemusí v každom ohľade súvisieť s myslou: „Zdá sa, že Turingove pojmy inteligencia a inteligentný nesúvisia explicitne s myslením alebo myslou. Inteligencia skôr korešponduje s učením alebo pamäťou.“[\[45\]](#) To by nasvedčovalo, že Turingov návrh o inteligentných učiacich sa strojoch nemusel byť úplnou kópiou ľudskej mysli, ale iba inšpiráciou kognitívnych schopností pre stroje nevyužívajúce iba komputáciu. Nejde síce o presný obraz strojovej inteligencie, minimálne je však možné tieto predpoklady považovať za prístup k strojovej inteligencii, ktorý nemusí nutne súvisieť s myslou, a pritom je porovnateľný so súčasným DL.

Záver

Ukázalo sa, že možno identifikovať Turingov vplyv na súčasné učiace sa systémy. Avšak je dôležité dodať, že ide len čiastočné implikácie a podobnosti s DL.

Turingov stroj je z technického hľadiska teoretickým najvýkonnejším počítačom, na ktorom možno dokazovať Turingovu úplnosť. V súvislosti s DL systémami sa však javí ako nedostačujúci. Na oboch prípadoch (Super-Turingov model, Turingov neurónový stroj) bolo možné vidieť, že disponujú vlastnosťami ako adaptácia či interakcia, na ktoré je iba komputácia nedostatočná.

Význam Turingových konekcionistických návrhov sa nachádza hlavne v modifikovateľnosti a organizovateľnosti neorganizovaných strojov. Na základe modifikovateľnosti bolo možné ukázať, že interferencia a samo-modifikovateľnosť je porovnateľná so súčasnými DL metódami.

Organizovateľnosť priamo inklinuje k učeniu, čo korešponduje s učiacimi sa strojmi.

Z konekcionistického hľadiska je organizovanie neorganizovaného detského mozgu prijateľné ako koncept, z ktorého Turing vychádzal. Na druhej strane je takýto koncept potrebné vnímať skôr ako analógiu s DL než priamu súvislosť. Súčasne je problematická univerzálnosť, keďže sa organizovaním stroj univerzálnym nestane, najmä nie v podobe, v akej by si Turing predstavoval. Keďže i neurónové siete sú neorganizované iba pred tréningom, keďže sa po natrénovaní stanú uzavretými.

V prípade učenia je detská myseľ pre Turinga taktiež dôležitým predpokladom. Hlavnými dôvodmi je prázdna pamäť (nepopísaný zošit) a málo počiatočných mechanizmov. Problémom však môže byť implementácia mechanizmov, pretože súčasné vedecké poznatky úplne nekorešponujú s jeho návrhmi. Zároveň sa ukazuje ako irelevantný i výber detskej mysle, keďže i dospelá myseľ môže disponovať rovnakými atribútmi ako detská. Na problém sa bolo možné pozrieť aj v súvislosti s DL, pričom nebolo nutné brať ohľad na konkrétny druh mysle. Metódu RL bolo možné priamo pripodobniť k zložkám, ktorými by mal podľa Turinga disponovať učiaci sa stroj. Na základe toho bolo možné objasniť vzdelávanie, schopnosť interakcie i získavanie skúseností, hoci takýto prístup skrýva obmedzenie v nemožnosti dosiahnuť univerzálnosť. Turing si to však pravdepodobne uvedomoval, pretože v spojení s učiacimi sa strojmi uvažoval aj o úlohách, ktoré by vykonávali.

V spojitosti so všetkými rozoberanými časťami a Turingovým predpokladom „o inteligentných strojoch s odklonom v správaní“ možno usúdiť, že s výnimkou Turingovho stroja sú predpoklady o učiacich sa strojoch porovnateľné so súčasným DL. I keď niektoré návrhy – najmä o detskej mysli a mozgu – vyznievajú vágne, niet pochyb, že načrtnol základy UI založenej na konekcionizme, ktorá vďaka svojim schopnostiam môže priniesť zásadnejší pokrok v strojovej inteligencii ako klasická UI.

Literatúra

ABRAMSON, D.: Turing's Responses to Two Objections. In: *Minds & Machines*, 18, 2008 s. 147-167. DOI: <https://doi.org/10.1007/s11023-008-9094-6>.

ANGUERA, A. - LARA, J.A. - LIZCANO, D. et al. Turing: The Great Unknown. In: *Found Sci* 25, 2019, s. 1203-1225. DOI: <https://doi.org/10.1007/s10699-019-09596-6>.

COPELAND, B. J. - PROUDFOOT, D.: On Alan Turing's anticipation of connectionism. In: *Synthese: An International Journal for Epistemology, Methodology and Philosophy of Science*, 108, 1996, pp. 361-377. DOI: <https://doi.org/10.1007/BF00413694>.

COPELAND, B. J. - PROUDFOOT, D.: What Turing Did after He Invented the Universal Turing Machine. In: *Journal of Logic, Language, and Information*, vol. 9, 2000, pp. 491-509. DOI: <https://doi.org/10.1023/A:1008371426608>.

DAVIS, M.: Turing's Vision and Deep Learning. In: Manea, F., Miller, R., Nowotka, D. (eds.): *Sailing Routes in the World of Computation. CiE 2018. Lecture Notes in Computer Science*, vol. 10936, 2018, pp. 146-155. DOI: https://doi.org/10.1007/978-3-319-94418-0_15.

GRAVES, A. - WAYNE, G. - DANIHELKA, I.: Neural turing machines. In: *arXiv preprint arXiv:1410.5401*, 2014, p. 26. DOI: <https://doi.org/10.48550/arXiv.1410.5401>.

GROSZ, B.: What Question Would Turing Pose Today? In: *AI Magazine*, vol. 33, no. 4, 2012, pp.

73-81. DOI: <https://doi.org/10.1609/aimag.v33i4.2441>.

HALL, J. S.: Self-improving AI: an Analysis. In: *Minds & Machines* 17, 2007, pp. 249-259. DOI: <https://doi.org/10.1007/s11023-007-9065-3>

HERNÁNDEZ-ORALLO, J. - DOWE, D. L.: On Potential Cognitive Abilities in the Machine Kingdom. In: *Minds & Machines* 23, 2013, pp. 179-210. DOI: <https://doi.org/10.1007/s11023-012-9299-6>.

KUGEL, P.: Computing Machines Can't Be Intelligent (...and Turing Said So). In: *Minds and Machines*, 12, 2002, pp. 563-579. DOI: <https://doi.org/10.1023/A:1021150928258>.

LYRE, H.: The State Space of Artificial Intelligence. In: *Minds & Machines*, 30, 2020, pp. 325-347. DOI: <https://doi.org/10.1007/s11023-020-09538-3>.

MCCULLOCH, W. S. - W. PITTS: A Logical Calculus of the Ideas Immanent in Nervous Activity. In: *Bulletin of Mathematical Biophysics* 5, 1943, pp. 115-133. DOI: <https://doi.org/10.1007/BF02478259>.

MILKOWSKI, M.: Turing's Conceptual Engineering. In: *Philosophies* 7, 69, 2022, p. 20. DOI: <https://doi.org/10.3390/philosophies7030069>.

RUSSEL, S.: *Jako člověk. Umělá inteligence a problém jejího ovládní*. Praha: Dokořán, s. r. o., 2021. ISBN: 978-80-7675-043-2.

SIEGELMANN, H. T.: Turing on Super-Turing and adaptivity. In: *Progress in Biophysics and Molecular Biology*, vol. 113, no. 1, 2013, pp. 117-126. DOI: <https://doi.org/10.1016/j.pbiomolbio.2013.03.013>.

TEUSCHER, CH.: *Turing's Connectionism: An Investigation of Neural Network Architectures*. London: Springer Science & Business Media, 2002. ISBN: 978-1-85233-475-8.

TURING, A.: Computing Machinery and Intelligence. In: *Mind*, vol. 59, no. 236, 1950, pp. 433-460. DOI: <https://doi.org/10.1093/mind/LIX.236.433>.

TURING, A.: Intelligent machinery. In: *National Physical Laboratory*. Edinburgh - Scotland, 1948, pp. 107-127. URL: <https://weightagnostic.github.io/papers/turing1948.pdf>.

P o z n á m k y

- [1] Príspevok vznikol vďaka podpore VEGA 2/0163/22 - *Literatúra v bioetike a bioetika v literatúre*.
- [2] Automatic Computing Engine - viacúčelový digitálny počítač navrhnutý Turingom.
- [3] TURING, A.: Computing Machinery and Intelligence. In: *Mind*, vol. 59, no. 236, 1950, pp. 433 - 460.
- [4] TURING, A.: Intelligent machinery. In *National Physical Laboratory*. Edinburgh - Scotland, 1948, pp. 107 - 127.
- [5] COPELAND, B. J. - PROUDFOOT, D.: What Turing Did after He Invented the Universal Turing Machine. In: *Journal of Logic, Language, and Information*, vol. 9, 2000, p. 491.
- [6] ABRAMSON, D.: Turing's Responses to Two Objections. In: *Minds & Machines*, 18, 2008 pp. 147 - 148.
- [7] Strojové učenie (machine learning, ML) je oblasť UI založená na učiacich sa algoritmoch. Hlboké učenie (deep learning, DL) je podkategóriou strojového učenia využívajúceho viacvrstvové neurónové siete.
- [8] TURING, A. Computing Machinery and Intelligence. In: *Mind*, vol. 59, no. 236, 1950, pp. 459.
- [9] Ibid., p. 437.
- [10] RUSSEL, S.: *Jako člověk Umělá inteligence a problém jejího ovládní*. Praha: Dokořán, s. r. o., 2021, p. 36.
- [11] Počítač, programovací jazyk či iný systém, ktorý dokáže vypočítať toľko ako Turingov stroj. Inak povedané počítač spĺňa Turingovu úplnosť, ak má dostatočnú pamäť na výpočet problému, na ktorý je určený.
- [12] HERNÁNDEZ-ORALLO, J. DOWE, D. L.: On Potential Cognitive Abilities in the Machine Kingdom. In: *Minds & Machines* 23, 2013, p. 196.
- [13] SIEGELMANN, H. T.: Turing on Super-Turing and adaptivity. In: *Progress in Biophysics and Molecular Biology*, vol. 113, no. 1, 2013 pp. 117 - 126.
- [14] GRAVES, A. - WAYNE, G. - DANIHELKA, I.: Neural turing machines. In: *arXiv preprint*

arXiv:1410.5401, 201, 2014, pp. 26.

[15] Konkrétne sa pokúsila dokázať, že ARNN siete (typ neurónových sietí) môžu byť matematicky výkonnejšie ako Turingov stroj. „Turingov stroj sa považuje za najvýkonnejšiu formu počítačových výpočtov. Pokúsila som sa zistiť, či je ARNN sieť schopná vypočítať viac ako Turingov stroj. Bola som úplne prekvapená, keď som zistila, že ARNN sieť je matematicky silnejšia. Na základe toho som aj pomenovala tento model Super-Turing.“ SIEGELMANN, H. T.: Turing on Super-Turing and adaptivity. In: *Progress in Biophysics and Molecular Biology*, vol. 113, no. 1, 2013, p. 121. Ide o matematický dôkaz, že je možné na súčasnej technológii vypočítať viac, ako umožňuje Turingov stroj.

[16] Ibid., p. 120.

[17] GRAVES, A. - WAYNE, G. - DANIHELKA, I.: Neural turing machines. In: *arXiv preprint arXiv:1410.5401*, 201, 2014, p. 22.

[18] Presnejšie je pamäť v neurónovom Turingovom stroji schopná pracovať s viacerými premennými súčasne v porovnaní so štandardným digitálnym počítačom alebo Turingovým strojom, ktoré dokážu pracovať iba s jednou. Pozri aj p. 5.

[19] KUGEL, P.: Computing Machines Can't Be Intelligent (...and Turing Said So). In: *Minds and Machines* 12, 2002, p. 565.

[20] Za úplne prvú prácu sa považuje MCCULLOCH, W. S. - W. PITTS: A Logical Calculus of the Ideas Immanent in Nervous Activity. In: *Bulletin of Mathematical Biophysics* 5, 1943, 1 15 - 133. McCulloch však priznal, že sa nechal taktiež inšpirovať staršími Turingovými prácami.

[21] COPELAND, B. J. - PROUDFOOT, D.: On Alan Turing's anticipation of connectionism. In: *Synthese: An International Journal for Epistemology, Methodology and Philosophy of Science*, 108, 1996, p. 361.

[22] TURING, A.: Intelligent machinery. In: *National Physical Laboratory*. Edinburgh - Scotland 1948, p. 117.

[23] To znamená, že môžu fungovať aj ako program v počítači.

[24] TEUSCHER, CH.: *Turing's Connectionism: An Investigation of Neural Network Architectures*. Londýn: Springer Science & Business Media, 2002, p. 17.

[25] TURING, A.: Intelligent machinery. In: *National Physical Laboratory*. Edinburgh - Scotland, 1948. pp. 115 - 116.

[26] Podobnosť SL a klasickej UI možno vidieť v tom, že u SL je pri dostatočnom množstve dát a natrénovaní výsledok vysoko predpokladaný.

[27] Ibid., p. 115.

[28] Pri nedostačnom množstve vzorových dát by mohlo DL vyhodnotiť úplný nezmysel. Napr. že pes je kôň.

[29] Schopnosť samostatne organizovať vstupné dáta a zisťovať v nich nepredpokladané vlastnosti a vzory. Napr. pozri LYRE, H.: The State Space of Artificial Intelligence. In: *Minds & Machines*, 30, 2020 pp. 325 - 347.

[30] ANGUERA, A. - LARA, J.A. - LIZCANO, D. et al.: Turing: The Great Unknown. In: *Found Sci* 25, 2019, p. 1221.

[31] TEUSCHER, CH.: *Turing's Connectionism: An Investigation of Neural Network Architectures*. Londýn: Springer Science & Business Media, 2002, p. 83.

[32] Či už ide o siete ANN - umelé neurónové siete, RNN - Rekurentné neurónové siete alebo CNN - koevolučné neurónové siete.

[33] TURING, A.: Intelligent machinery. In: *National Physical Laboratory*. Edinburgh - Scotland, 1948, p. 111.

[34] Anguera et al. ukázali, že Turingove modifikovateľné siete typu B sú v tomto podobné. „Pri modeli sietí B sa tréningom zničia nepotrebné spojenia, dochádzajú modifikácie spojené s režimom ‚prerušenia‘. Recipročne sa mení modifikátor z režimu ‚prerušenie‘ do režimu ‚prechod‘ a tým účinne vytvorí spojenie. Tento selektívny výber a uvoľnenie spojení mení pôvodne náhodnú sieť na organizovanú sieť na vykonávanie konkrétnej úlohy.“ ANGUERA, A. - LARA, J.A. - LIZCANO, D. et

al.: Turing: The Great Unknown. In: *Found Sci* 25, 2019, p. 1221.

[35] Technicky je síce treba vopred vybrať metódu a typ neurónovej siete, vo všeobecnosti to tak však platí.

[36] TURING, A.: Computing Machinery and Intelligence. In: *Mind*, vol. 59, no. 236, 1950, p. 456.

[37] GROSZ, B.: What Question Would Turing Pose Today? In: *AI Magazine*, vol. 33, no. 4, 2012, p. 77.

[38] HALL, J. S.: Self-improving AI: an Analysis. In: *Minds & Machines* 17, 2007, p. 250.

[39] TURING, A.: Computing Machinery and Intelligence. In: *Mind*, vol. 59, no. 236, 1950, p. 456.

[40] TEUSCHER, CH.: *Turing's Connectionism: An Investigation of Neural Network Architectures*. London: Springer Science & Business Media, 2002, p. 118.

[41] TURING, A.: Computing Machinery and Intelligence. In: *Mind*, vol. 59, no. 236, 1950, p. 455.

[42] HERNÁNDEZ-ORALLO, J. - DOWE, D. L.: On Potential Cognitive Abilities in the Machine Kingdom. In: *Minds & Machines* 23, 2013, p. 199.

[43] TURING, A.: Computing Machinery and Intelligence. In: *Mind*, vol. 59, no. 236, 1950, p. 460.

[44] Pozri DAVIS, M.: Turing's Vision and Deep Learning. In: Manea, F., Miller, R., Nowotka, D. (eds.): *Sailing Routes in the World of Computation. CiE 2018. Lecture Notes in Computer Science*, vol. 10936, 2018, p. 147.

[45] MIŁKOWSKI, M.: Turing's Conceptual Engineering. In: *Philosophies* 7, 69, 2022, p. 1214.

Mgr. Ondrej Hriadel

Katedra filozofie a aplikovanej filozofie

Filozofická fakulta Univerzity sv. Cyrila a Metoda v Trnave

Nám. J. Herdu 2

917 01 Trnava

email: ondro.hriadel@gmail.com

ORCID: 0000-0001-6490-455X