



Srovnání umělé a biologické inteligence – filosofická perspektiva

Bakalářská práce

Studijní program:

B6101 Filozofie

Studijní obor:

Filozofie humanitních věd

Autor práce:

Lucie Firmanová

Vedoucí práce:

Mgr. Vít Bartoš, Ph.D.
Katedra filosofie





Zadání bakalářské práce

Srovnání umělé a biologické intelligence – filosofická perspektiva

Jméno a příjmení: **Lucie Firmanová**
Osobní číslo: P16000154
Studijní program: B6101 Filozofie
Studijní obor: Filozofie humanitních věd
Zadávající katedra: Katedra filosofie
Akademický rok: **2017/2018**

Zásady pro vypracování:

Práce se zaměří na vysvětlení a porovnání principů, jimiž se řídí biologické mysli a mysli umělé (lidmi vytvořené), přičemž téma bude pojednáno z obecné filosofické perspektivy. Východiskem práce budou základní intuice o rozdílnosti mezi umělou a lidskou inteligencí Johna Searlea. Studentka se bude řídit metodickými pokyny vedoucího práce a bude pracovat samostatně.

Rozsah grafických prací:
Rozsah pracovní zprávy:
Forma zpracování práce:
Jazyk práce:

tištěná/elektronická
Čeština



Seznam odborné literatury:

- HAVEL, Ivan M. Úvod. In: MAŘÍK, Vladimír a kolektiv. Umělá inteligence (6). Praha: Academia, 2013. ISBN 978-80-200-2267-9
- HAVEL, Ivan M. Přirozené a umělé myšlení jako filozofický problém. In: MAŘÍK, Vladimír a kolektiv. Umělá inteligence (3). Praha: Academia, 2001. ISBN 80-200-0472-6.
- ROMPORTL, Jan. Přirozenost umělé inteligence. In: BENEDIKTOVÁ, Marie et al. Člověk v nových světech. Plzeň: Vydavatelství Západočeské univerzity v Plzni, 2012. ISBN 978-80-261-0068-3.
- SEARLE, John. Minds, brains, and programs. The Behavioral and Brain Sciences, 1980, 3(3), pp. 417-457.
- SEARLE, John R. Mysl, mozek a věda. Přeložil Marek NEKULA. Praha: Mladá fronta, 1994. ISBN 80-204-0509-7.
- TVRDÝ, Filip. Turingův test: filozofické aspekty umělé inteligence. Praha: Togga, 2014. ISBN 978-80-7476-043-3.
- HAVEL, Ivan M. Úvod. In: MAŘÍK, Vladimír a kolektiv. Umělá inteligence (1). Praha: Academia, 1993. ISBN 80-200-0496-3

Vedoucí práce:

Mgr. Vít Bartoš, Ph.D.
Katedra filosofie

Datum zadání práce:

20. dubna 2018

Předpokládaný termín odevzdání:

30. dubna 2019

prof. RNDr. Jan Pícek, CSc.
děkan

L.S.

doc. PhDr. David Václavík, Ph.D.
vedoucí katedry

Prohlášení

Prohlašuji, že svou bakalářskou práci jsem vypracovala samostatně jako původní dílo s použitím uvedené literatury a na základě konzultací s vedoucím mé bakalářské práce a konzultantem.

Jsem si vědoma toho, že na mou bakalářskou práci se plně vztahuje zákon č. 121/2000 Sb., o právu autorském, zejména § 60 – školní dílo.

Beru na vědomí, že Technická univerzita v Liberci nezasahuje do mých autorských práv užitím mé bakalářské práce pro vnitřní potřebu Technické univerzity v Liberci.

Užiji-li bakalářskou práci nebo poskytnu-li licenci k jejímu využití, jsem si vědoma povinnosti informovat o této skutečnosti Technickou univerzitu v Liberci; v tomto případě má Technická univerzita v Liberci právo ode mne požadovat úhradu nákladů, které vynaložila na vytvoření díla, až do jejich skutečné výše.

Současně čestně prohlašuji, že text elektronické podoby práce vložený do IS/STAG se shoduje s textem tištěné podoby práce.

Beru na vědomí, že má bakalářská práce bude zveřejněna Technickou univerzitou v Liberci v souladu s § 47b zákona č. 111/1998 Sb., o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších předpisů.

Jsem si vědoma následků, které podle zákona o vysokých školách mohou vyplývat z porušení tohoto prohlášení.

12. května 2020

Lucie Firmanová

Anotace

Bakalářská práce „Srovnání umělé a biologické inteligence – filosofická perspektiva“ se zabývá pojmem inteligence a porovnává inteligenci umělou a biologickou.

V první kapitole se budu zabývat pojmem inteligence. Nejprve se pokusím tuto kategorii definovat. Dále se budu věnovat inteligenci lidské, zvířecí a strojové. Rovněž se pokusím přiblížit pojem umělá inteligence a nastínit, kdy se tento fenomén poprvé objevil a jaká jsou jeho úskalí.

Druhá kapitola se bude věnovat Alanu Turingovi, autorovi podstatné části principů, na jejichž základě pracují moderní počítače. Cílem této kapitoly bude nastínit průběh a význam Turingova testu, imitační hry a vysvětlit problém definice pojmů „myslet“ a „stroj“. Pozornost bude též věnována prvním kritikám Turingovy koncepce a námitkám vůči jeho testům.

Třetí kapitola bude věnována americkému filozofovi J. R. Searlovi a jeho známému argumentu čínského pokoje. Nastíněn bude též problém mysli a těla (tzv. dualismus) a pojednána otázka, zda mohou počítače myslet.

Čtvrtá a zároveň poslední kapitola bude konfrontovat dva přístupy – kybernetický funkcionalismus a biologický naturalismus – a podrobí je kritice.

Klíčová slova

Umělá inteligence, lidská inteligence, Turingův test, filosofie mysli, Argument čínského pokoje, vědomí, mysl,

Annotation

Bachelor thesis "Comparison of artificial and biological intelligence - philosophical perspective" deals with the concept of intelligence and the comparison between artificial and biological intelligence.

The first chapter will cover the concept of intelligence. First, I will try to define the term. Next, I will focus on the human, animal and machine intelligence. Then I will try to explain the term artificial intelligence and describe when did the phenomenon first appear and what are its pitfalls.

The second chapter will deal with Alan Turing, who devised a substantial part of the principles behind modern computers. The aim of this chapter is to explain the content and meaning of the Turing test, imitation games and the problem of defining of the terms "think" and "machine". There will also appear some of the critical voices to the Turing's concept and his experiments.

The third chapter will be devoted to an American philosopher J. R. Searle and his well-known argument of the Chinese room. The problem of mind and body (dualism) will also be covered there, as well as the question of whether computers are able to think.

The fourth and final chapter will deal confrontate the concepts of cybernetic functionalism and biological naturalism and criticize them. Furthermore, I will mention a Czech scientist I. M. Havel, who also deals with artificial intelligence.

Keywords

Artificial Intelligence, human Intelligence, Turing Test, mind philosophy, Chinese Room Argument, consciousness, mind

Poděkování

Tímto bych chtěla velmi poděkovat svému profesorovi a vedoucímu práce Mgr. Vítu Bartošovi, Ph.D., za odborné vedení, jeho přínosné připomínky, kritické rady a toleranci při vytváření této bakalářské práce.

Obsah

Úvod.....	9
1. Intelligence.....	15
Intelligence lidská	16
Intelligence zvířat.....	17
Intelligence/ mysl strojů a počítačů.....	21
2. Alan Turing.....	23
Esej Computing Machinery and Intelligence.....	23
Turingovy námitky.....	25
Předchůdce Turingova testu.....	30
3. John Rogers Searle.....	31
Problém mysli a těla.....	31
Mohou počítače myslet?	34
Kognitivní věda.....	35
Argument čínského pokoje	37
4. Kybernetický funkcionalismus versus biologický naturalismus.....	43
Kybernetický funkcionalismus.....	44
Biologický naturalismus	46
5. Závěr	49
Seznam použité literatury:.....	52
Seznam použitých webových stránek:	54

Úvod

Téma bakalářské práce „Srovnání umělé a biologické intelligence – filosofická perspektiva“ jsem si vybrala na základě svého zájmu o umělou inteligenci a převážně o část tohoto oboru, která se zabývá právě srovnáním umělé a biologické intelligence. Jelikož jsem studentkou Přírodovědně-humanitní a pedagogické fakulty, považuji za korektní psát převážně jen o filozofických otázkách spojených s umělou inteligencí. Svou práci zaměřím na definici umělé intelligence z pohledu zakladatele Alana Turinga a na jeho imitační hry. Jako kontrast chci podat zprávu o skeptickém názoru Johna R. Searla, jehož myšlenkový experiment čínského pokoje je pokládán za protiargument k Turingovu testu. Práce se dále zaměří na analýzu a porovnání principů, na jejichž základě pracují biologické mysli a mysli umělé (lidmi vytvořené).

Jedním z klíčových témat filosofie napříč jejím vývojem je problematika mysli a těla a vztahu mezi nimi. Ačkoliv se s tzv. psychofyzickým problémem setkáváme už v antické filosofii, nejvíce ho rozpracovala novověká filosofie a poté ve 20. stol. filosofie mysli. V diskusi o mind-body problému dlouho přetrvával dualismus, dnes se dává přednost spíše monistickým teoriím mysli a těla, příp. umírněnému dualismu (např. u Searla).¹

Zatímco u Descarta je mysl úzce propojena s racionalitou, filosofie mysli pod tímto pojmem chápe soubor psychických vlastností, procesů a stavů. Mysl tedy zahrnuje rovněž afekty, emoce, volní procesy apod. S rozvojem techniky vyvstává otázka, zda mysl můžeme připisovat nejen lidem, ale rovněž neživým strojům lidmi vytvořeným. Srovnání člověka s precizně fungujícím strojem přináší již novověký mechanicismus (např. de La Mettrie). K plnému rozvinutí úvah o schopnosti uměle vytvořených bytostí myslet však logicky došlo teprve v době, kdy již technika byla na takové úrovni, aby bylo možné tyto „myslící“ stroje vytvořit.

Téma rozpracoval britský matematik Alan Turing, který patří mezi významné myslitele 20. století a který se podílel na rozvoji několika vědeckých odvětví (kryptografie, kybernetiky, matematiky či filosofie mysli).

1 Polák, 2013, s. 40 – 41.

Do povědomí veřejnosti se nejvíce zapsal svým podílem na prolomení kódu šifrovacího stroje Enigma, jehož prostřednictvím spolu komunikovaly německé válečné ponorky během druhé světové války.²

Roku 1950 Turing publikoval filosofickou esej *Computing Machinery and Intelligence*, kde si jako první položil otázku, zda mohou počítače myslet. Zde se také objevuje první zmínka o Turingovu konceptu α -stroje (dnes známý jako Turingův stroj), kterým se budu zabývat v další kapitole.

Ačkoliv můžeme přímo vnímat pouze své vlastní mentální procesy,³ mysl automaticky připisujeme rovněž ostatním lidem. Tento fakt lze vysvětlit hned ze 3 filosofických stanovisek:⁴

1. **Řešení z pozice zdravého rozumu.** Přítomnost mysli u ostatních lidí zpravidla považujeme za truismus, tedy tvrzení, které je všeobecně přijímáno jako pravdivé, a není tudíž nutné jeho správnost dokazovat či o něm polemizovat. Taková skutečnost je natolik evidentní, že není třeba ji vůbec ani zmiňovat.
2. **Pragmatické řešení.** Není výhodné, abychom pochybovali o přítomnosti mysli u ostatních lidí a chovali se k nim tak, jako by mysl neměli. V takovém případě by mohlo být rovněž těmito lidmi pochybováno o existenci naší vlastní mysli, přesně tedy „*V běžném životě je i ten nejzatvrzelejší skeptik k nerozeznání od ostatních smrtelníků*“.⁵ Ve vlastním zájmu tedy přijímáme nedokazatelný předpoklad, že všichni lidé jsou myslící bytosti a v souladu s tímto předpokladem s lidmi také jednáme.

2 V roce 2014 byl uveden film *Kód Enigmy* (v orig. *The Imitation Game*), který dokumentuje příběh Alana Turinga, profesora matematiky a kryptoanalytiky, který se spolu s týmem podílel na prolomení kódu legendárního německého šifrovacího stroje Enigma. Tím pomohl zkrátit 2. světovou válku a zachránit tisíce lidských životů. (zdroj: <https://www.csfd.cz/film/283747-kod-enigmy/prehled/>)

3 Pojmem mentální entity rozumějme veškeré psychické stavy a události. Zahrňme sem tedy počítky, kognitivní stavy, emoce, vjemy (percepce), kvazi-vjemové stavy a konativní stavy. Polák, 2015, s. 31.

4 Tvrdý, 2014, s. 8 – 9.

5 HUME, David, 1996. *Zkoumání o lidském rozumu*. Přel. Josef MOURAL. Praha: Svoboda, s. 216

3. **Řešení na základě analogie.** Na základě námi přijímaných vědeckých poznatků i vlastního pozorování jsme si vědomi toho, že ostatní lidé mají (až na drobné výjimky) dosti podobnou anatomickou stavbu jako my sami. Týká se to rovněž nervové soustavy. S jistotou víme, že my sami máme mysl, a připisujeme ji i bytostem s podobnou tělesnou a zejména nervovou stavbou, tedy všem lidem.

Ani jedno z těchto řešení však nelze uspokojivým způsobem použít, chceme-li připsat mysl rovněž bytostem, které nejsou lidmi. V případě počítačů, robotů a dalších umělých entit selhává zdravý rozum. Málokdo totiž takové přístroje považuje za skutečně inteligentní, a pokud ano, tak jen v případě, že jsou zapnuté. Mysl jim rovněž nelze připsat na základě analogie, nemají totiž žádnou nervovou soustavu a jejich stavba je celkově zcela odlišná od té naší. A co se týče pragmatického řešení, v našem zájmu je, aby myslícími entitami zůstali výhradně lidé (z různých důvodů, které nejsou pro tuto práci podstatné).⁶

Když se zamyslíme, tak nad každou lidskou činností si lze klást otázky, které přesahují samu tuto činnost. Touha po poznání sebe sama a svého konání je koneckonců lidskou přirozeností, povinností a darem. V souvislosti s umělou inteligencí vyvstává řada otázek, které si můžeme rozdělit do 5 kategorií – na otázky ontologické, epistemologické, metodologické, futurologické a etické:⁷

- **Ontologické otázky** se ptají po samé podstatě bytí toho, co zkoumáme, tedy v našem případě umělé inteligence. Zabývají se vztahem přírodního (či, chceme-li, přirozeného) a umělého, lidmi vytvořeného. Můžeme sem rovněž zařadit dlouhou historii diskusí o vztahu mysli a těla (*mind-body problem*) nebo zkoumání významu pojmů duše, mysl, vědomí apod.
- **Epistemologické otázky** mají za cíl zjistit, jak se lidský způsob poznávání a myšlení liší od umělého, strojového a zda jsou vůbec počítače schopny myslet. Zkoumají se rovněž způsoby, jakými se lze dozvídat o myšlení lidí a jiných entit, a porovnávat je.

6 Tvrđý, 2014, s. 12.

7 Havel, 2001, s. 17 – 18.

- **Metodologické otázky** vyvstávají v souvislosti se studiem metod v rámci jednotlivých oborů umělé inteligence nebo kognitivní vědy, jejich cílů a motivací, zásad zkoumání, postupů, pravidel pro sestavování hypotéz a provádění experimentů apod. Zařadit sem můžeme taktéž úvahy nad tím, které z nabízených paradigmat je v té které oblasti zkoumání vhodné přijmout.
- **Futurologické otázky** se snaží na základě logiky a znalosti historie, matematiky a umělé inteligence(dále jen UI) odhadnout, jaký vývoj lze v oblasti UI předpokládat do blízké i vzdálené budoucnosti, a především jaký dopad bude mít rozvoj umělé inteligence na lidstvo i planetu jako celek.
- **Etické otázky** se zabývají morálními aspekty umělé inteligence. Zamýšlejí se nad tím, zda je morálně správné rozvíjet umělou inteligenci, nahrazovat lidské subjekty v různých oblastech umělými (např. stroje v továrnách nebo samořídící auta, kolem nichž v současné době probíhá rozsáhlá diskuse) apod. Patří sem i obecný problém, zda a v kterou chvíli se má vědec řídit vlastním názorem či pocitem, že jeho práce vede ku prospěchu lidstva a světa vůbec, či k jeho zkáze.

Problematika zkoumání lidské mysli

I. M. Havel⁸ se domnívá, že pro zkoumání filosofických aspektů umělé inteligence je nezbytné nejprve zkoumat lidské myšlení, nikoliv pouze provádět počítačová modelování. Samozřejmě, pokud se snažíme napodobit lidskou mysl a vložit ji do počítačů a robotů, musíme mít dostatek informací o tom, jak funguje. Zkoumání přirozené mysli můžeme provádět třemi způsoby:⁹

- **Introspektivní metoda** – v psychologii byla dlouho považována za nevědeckou, nespolehlivou a nic nevypovídající, ostatně s tímto pohledem se setkáváme i dnes. Je však třeba si přiznat, že tato metoda je velmi jednoduchá, a přitom přináší množství informací. Je dostupná každému, neboť každý člověk je schopen nahlížet své vlastní mentální entity, přemýšlet a vypovídat o nich. Výsledky introspektivního zkoumání jsou tedy nutně subjektivní – o mentálních stavech a procesech může přímo hovořit pouze jedinec, v němž se tyto odehrávají, ostatní lidé o nich mohou vypovídat jen zprostředkovaně. Zároveň však lze předpokládat, že mají podobnou zkušenost, a proto jsou schopni jim porozumět. Vzhledem k nedostatkům introspekce patrně nelze lidskou mysl zkoumat výhradně tímto způsobem, je ale vhodné jím doplnit poznatky získané jinými metodami.
- **Biologické zkoumání** nemá k dispozici mentální stavy izolované od prostředí, a proto se věnuje buď vnějším projevům těchto stavů (tedy chování), anebo orgánů, v nichž se odehrávají (tedy mozku). Tímto způsobem zkoumají lidské myšlení neurobiologie, psychologie a neuropsychologie, která se snaží propojit poznatky obou předchozích.
- **Umělé modelování** je nepřímou metodou zkoumání lidské mysli, spočívá totiž ve zkoumání uměle vytvořených modelů. Teze, že počítač může sloužit jako důvěryhodný model lidské mysli, významně ovlivnila kybernetický funkcionalismus (kap. 4.1). Vystupují zde dva druhy modelů – věrný a metaforický. Věrný model je zpravidla fyzický (může být ale též matematický) model, který má co nejvěrněji napodobovat svůj vzor, a to v co nejvíce možných aspektech.

8 I. M. Havel je český vědecký pracovník, který se zabývá umělou inteligencí.

9 Tamtéž, s. 22 – 27.

U metaforických modelů naopak není na závadu, když se v něčem od svého vzoru liší, i v takovém případě totiž získáváme důležité poznatky. Metaforický model může mít fyzickou nebo abstraktní (teoretickou) povahu či může být softwarovou simulací.

1. Intelligence

Pojem intelligence lze definovat mnoha různými způsoby, nejčastěji však odkazuje ke schopnosti chápání, logického a kritického myšlení, k řešení problémů, učení, plánování, kreativitě a schopnosti vyjadřovat emoce. Intelligence byla chápána jako soubor určitých měřitelných schopností, a proto byly v 1. polovině 20. století vytvářeny testy intelligence (vůbec první test vytvořil již Francis Galton v r. 1883¹⁰). V 80. letech však pracovník Harvardské univerzity Howard Gardner vyslovuje myšlenku, že inteligencí je vícero druhů. Dnes je obecně uznáváno 9 druhů intelligence, které mají společně pokrývat všechny sféry lidského života (přírodovědná intelligence, hudební i., logicko - matematická i., existenciální i., interpersonální i., tělesně-pohybová i., jazyková i., intrapersonální i., prostorová i.).¹¹

Samotné slovo intelligence pochází z latinského *inter-legere*, což v překladu znamená rozlišovat, poznávat, chápat. Ve středověké filosofii a teologii bylo časté používání termínu *intellectus*, což je jakási obdoba řeckého *nús* (νοῦς) a má spíše metafyzický a epistemologický význam. Naproti tomu pojem intelligence (z *intellectus* odvozený) je užíván moderní psychologií.

Časopis *Journal of Educational Psychology* v roce 1921 požádal 14 významných psychologů, aby definovali pojem intelligence. Výsledkem bylo 14 různých definic, z nichž byla nakonec vytvořena tato: „*Intelligence je schopnost myšlení a učení se ze zkušeností. Schopnost přizpůsobit se okolnímu prostředí a adaptace, která se projevuje intelektovým výkonem,*“.¹²

V průběhu dějin se zrodila řada různých teorií intelligence, které můžeme rozdělit do 4 hlavních směrů – sociální, psychometrické, kognitivní a další. Pro tuto práci však jejich rozlišení není podstatné a proto přejdeme k rozdílu mezi lidskou, zvířecí a strojovou inteligencí.

10 Kassin, 2007.

11 <https://examinedexistence.com/the-nine-different-types-of-intelligence/>

12 <https://vesmir.cz/cz/on-line-clanky/2015/06/intelligence-co-vlastne-je.html>

Intelligence lidská

Jako první a možná nejdůležitější uvedu inteligenci lidskou, kterou má každý člověk na planetě. Jedním ze základních a důležitých zdrojů individuálních rozdílů u lidských bytostí jsou právě intelligence a osobnost. Jak jsem již zmínila, definovat takto členitou lidskou vlastnost je velmi náročné. Definice je i přes snahy řady významných psychologů zatím nedokonalá, lze ale předpokládat její další vývoj. Za všech definic zmíním alespoň tyto:

William Stern: *„Intelligence je všeobecná schopnost individua vědomě orientovat vlastní myšlení na nové požadavky, je to všeobecná duchovní schopnost přizpůsobit se novým životním úkolům a podmínkám,“*¹³

David Wechsler: *„Intelligence je vnitřně členitá a zároveň globální schopnost individua účelně jednat, rozumně myslet a efektivně se vyrovnávat se svým okolím,“*¹⁴

Joy Paul Guilford: *„Intelligence je schopnost zpracovávat informace. Informacemi je třeba chápat všechny dojmy, které člověk vnímá,“*¹⁵

Není pochyb, že intelligence sestává z mnoha různých vlastností a dovedností. Můžeme mezi ně zařadit schopnost jedince nějakým způsobem se orientovat ve světě a uvědomovat si sám sebe jako nepřetržitě existující bytost. Pomocí rozumu kreativně řešit různé problémy, s nimiž se potýká, vyhledávat cesty k co nejefektivnějšímu dosažení svých cílů atd. Obecně tedy inteligenci připisujeme těm jedincům, u kterých můžeme pozorovat přítomnost vyšších kognitivních funkcí.

13 http://casopis.mensa.cz/veda/intelligence_a_jeji_mereni.html

14 http://casopis.mensa.cz/veda/intelligence_a_jeji_mereni.html

15 http://casopis.mensa.cz/veda/intelligence_a_jeji_mereni.html

Intelligence zvířat

Otázkou, zda mají živočichové mysl, dokáží rozumět dění okolo sebe a svobodně se rozhodovat, anebo se naopak řídí jen instinkty a nemají vědomí, se zabývají filosofové již po dlouhá staletí. Živočichové jsou rovněž podrobováni vědeckým experimentům a pozorováním za účelem zjistit, zda mysl mají či nikoliv. Po dlouhou dobu dominovalo přesvědčení, že zvířata sice vnímají, ale nerozumějí a nechápou, nemluví a nejsou rozumná. Vzhledem tomu, že podle našich představ zvířatům chybí rozum, si nejsou vědoma své smrtelnosti, neznají úzkost, nevyjadřují emoce, nemohou hřešit a nemusí nést odpovědnost za své činy. Především novověcí filosofové často považovali zvířata za pouhé dobře naprogramované stroje a jakékoliv jejich konání, které bychom u člověka považovali za projev emocí, vysvětlovali na základě fyzikálních zákonů akce a reakce.

Zjistit, zda zvířata mají mysl, je neobyčejně obtížné. O existenci mysli ostatních lidí se totiž přesvědčujeme nejčastěji tak, že nám tito lidé sami vyprávějí o tom, co cítí, na co myslí, co prožívají. Živočichové nám nic z toho sdělit nemohou (jedině svým chováním, zde je však riziko chybné interpretace), a proto můžeme mít sklon domnívat se, že nejsou inteligentními bytostmi. Na druhou stranu však předpokládáme mysl u dětí, které si zatím ještě neosvojily jazyk.

Americký vědec Daniel Clement Dennett zastává názor, že jazyk není podmínkou pro existenci mysli: „*Mysl bez jazyka, prohlašuje tradice a zdravý rozum, nepochybně existuje.... Živočichové jiní než lidé – či alespoň někteří z nich – jistě mají mentální životy,*“.¹⁶ To, že jedinec nemluví (či dokonce ani nemůže mluvit, jelikož k tomu není biologicky vybaven), ještě neznamená, že nemá mysl – Dennett hovoří o tzv. nekomunikativních myslích. Ty se dozajista liší od myslí mluvících lidí, především v tom, že jejich obsah je nám zcela nepřístupný. V souvislosti s jejich nepoznatelností vyvstává problém jak rozlišit, kteří živočichové jsou myslící a kteří se pouze nechávají vést instinkty.¹⁷

16 Dennett, 1996, s. 20 – 21.

17 O takové rozlišení se snaží mj. australský filosof, známý bojovník za práva zvířat, Peter Singer. Rozeznává živočichy, kteří jsou osobami (tj. mají vlastnosti jako sebeuvědomování, schopnost chápat minulost a budoucnost, starost o druhé apod.), a živočichy, kteří jsou schopni cítit bolest, ale nemají mysl. Singer, 1993, s. 117.

U některých živočichů bylo prokázáno, že jsou schopni velmi specifickým způsobem reagovat na aktuální situaci, přičemž takové reakci musela předcházet myšlenková operace pracující s minulou zkušeností i očekáváním budoucnosti. Příkladem budiž šimpanz Figan, který při krmení spatřil banán visící na větvi nad hlavou výše postaveného člena tlupy. Aby si zajistil, že si tento samec banánu nevšimne, (nebo ho po něm nebude vyžadovat, pokud se pro něj nyní natáhne) odešel Figan o kus dál, tam si sedl a čekal. Jakmile okolo něj prošel zmíněný samec, vrátil se Figan pod strom, banán utrl a snědl.¹⁸ Rysy inteligentního chování byly pozorovány a podrobně zkoumány rovněž u delfínů a dalších kytovců, např. skupina delfínů v oblasti Laguna v Brazílii již po sto let spolupracuje s místními rybáři – nažene ryby do jejich sítí a plácnutím ocasem o hladinu dají rybářům znamení k vytažení sítí. Ti potom delfíny odmění částí úlovku.¹⁹ Příkladů zvířecí inteligence je mnoho, uveďme namátkou ještě schopnost počítání objektů u papouška šedého jménem Alex – ten v angličtině vyslovoval číslice, které odpovídaly počtu objektů, jež ze všech nabízených splňovaly určité kritérium, např. měly konkrétní barvu nebo tvar.²⁰ Těmto a dalším živočichům vyznačujícím se vědomím, sebeovládáním, používáním nástrojů, budováním vztahů s dalšími jedinci svého druhu (či dokonce jiných druhů), vědomím vlastní minulosti a budoucnosti, schopností abstraktního myšlení apod. tedy můžeme inteligenci téměř s jistotou připsat.

Není pochyb o tom, že za existencí inteligence stojí dlouhý a složitý vývoj. Dennett evoluci živočišných druhů, vedoucí ve výsledku k myslícím bytostem, přirovnává ke čtyřpatrové věži, kterou nazval „tvoř-a-zkoušej“.

18 Tamtéž, s. 116.

19 Simmonds, 2013, s. 43 – 44.

20 Morwitz, 2014, s. 575.

Dennettova čtyřpatrová věž

Dennettova věž je jakési metaforické vyjádření pro různé evoluční úrovně živých tvorů. Každé patro věže dává organismu schopnost nalézat lepší řešení situací a znamená pokrok ve schopnostech vědomí, intencionality a mysli.

1. **Darwinovští tvorové**, což jsou prvotní nejjednodušší organismy. Přežili jen ti jedinci, kteří byli vybaveni lepšími kombinacemi genů, tedy byli od přírody lépe uzpůsobeni k životu v daném prostředí.²¹
2. **Skinnerovští tvorové**, u nichž už má pro přežití význam jakási elementární forma učení. Mezi darwinovskými tvory se po určité době objevili jedinci, kteří měli v genetické výbavě posilovače zkušeností. Pokud se takový tvor v nějaké situaci zachoval tak, že byl pro něj výsledek příznivý (např. unikl nebezpečí, našel potravu, partnera apod.), posilovač způsobil, že v příští obdobné situaci měl organismus sklony zachovat se stejně jako v předchozím případě, tedy zopakovat osvědčené chování.²²
3. **Popperovští tvorové**, kteří se od skinnerovských liší schopností předběžného výběru z možností. Zatímco skinnerovský jedinec reaguje na prvotní situaci zcela náhodně, a teprve v případě úspěchu může na základě zkušenosti své příští chování volit, popperovský jedinec má dispozice k vyloučení některých možností předem. Tím se snižuje šance, že organismus na špatnou volbu doplatí životem. Aby byl takový předvýběr vůbec možný, musí mít v sobě jedinec nashromážděn dostatek informací o fungování vnějšího světa.²³
4. **Gregoryovští tvorové**, kteří při rozhodování nevyužívají jen vlastní zkušenosti, ale rovněž zkušenosti jiných příslušníků svého druhu (resp. skupiny), a to pomocí nástrojů (fyzických i mentálních, především jazyka). To vede k neustálému zlepšování života – jedincův život se zlepší, když je mu dán potřebný nástroj.

21 Dennet, 2004, s. 82.

22 Tamtéž, s. 83.

23 Tamtéž, s. 88 – 90.

On potom může tento nástroj dále zdokonalovat, čímž učiní život svých nástupců ještě lepším. Je zde tedy vyžadován život ve skupině a vzájemná spolupráce všech členů této komunity.²⁴

Dennettův evoluční výklad vysvětluje mysl a vědomí z hlediska evoluce. Během ní dochází k rozvoji intencionality a mysli. Díky tomu je pak každý živočich schopen vykázat vždy o krok vyspělejší stupeň intencionality než organismus vývojově mu předcházející. Je ale třeba zmínit, že bez níže stojících tvorů by ti výše postavení nemohli existovat. Jsou totiž tvořeni evolucí vždy na jejich základu a zároveň je v sobě obsahují – buňky, z nichž se skládá naše tělo, můžeme považovat za darwinovské jedince.

Podle Dennetta by bylo chybou považovat zvířata za nemyslicí. Ostatní mysli (tedy i mysli zvířecí) jsou pouze velmi odlišné od mysli lidské. Kupříkladu ani nám nejvíce podobní živočichové - šimpanzi, si nejsou schopni (alespoň dle Dennetta) určité věci vybavit nebo o nich vytvořit mentální představu. Připomeňme však, že mentální entity nekomunikativních myslí nám nejsou přístupné, a tedy nemůžeme s jistotou říci, zda předchozí tvrzení platí či nikoliv.

24 Tamtéž, s. 96 – 97.

Intelligence/ mysl strojů a počítačů

Dnešní počítače jsou schopny mnoha činností, které (kdybychom je pozorovali u živých tvorů) bychom označili za projev inteligence. Filozofové i matematici se již několik stovek let²⁵ přou o to, zda se skutečně jedná o potvrzení, že uměle vytvořené bytosti jsou schopny myslet jako člověk, či pouze o schopnost takových entit přesně se řídit pokyny, které jim byly dány programátory.

Jak jsme viděli v úvodu, mysl nelze strojům připsat ani na základě zdravého rozumu, ani z pragmatických důvodů a už vůbec ne na základě analogie s lidskou nervovou soustavou. Pokud bychom tedy chtěli zastávat názor, že roboti, počítače a další stroje jsou schopny lidského myšlení, museli bychom najít jiné argumenty. O to se nejvýznamněji pokusil anglický matematik, logik, výpočetní technik a filosof myslí Alan Turing ve své esejí *Computing Machinery and Intelligence*. Ještě před samotným rozbořením této esejí však ještě pár slov o umělé inteligenci jakožto vědním oboru.

Umělá inteligence

Umělá inteligence (artificial intelligence – AI) je vědní obor, který spojuje poznatky matematiky, fyziky, psychologie, biologie, neurověd, genetiky a dalších. Využívá je k sestavení umělých bytostí, které budou schopny napodobovat lidské inteligentní chování.

Přesná a všemi uznávaná definice umělé inteligence neexistuje, ačkoliv pokusů o ni můžeme najít stovky. Jmenujme např. vymezení Marvina Minského: „*Umělá inteligence je věda o vytváření strojů nebo systémů, které budou při řešení určitého úkolu používat takového postupu, který – kdyby ho dělal člověk – bychom považovali za projev jeho inteligence*,“.²⁶ Obvyčejné počítače při řešení problémů zkoušejí různé možnosti, až dospějí k té nejvhodnější. Naproti tomu inteligentní stroje mají být schopny předem na základě určitých znalostí rozhodnout, která řešení by v daném případě mohla vést k úspěchu, a ostatní vůbec nepoužít.²⁷

Umělá inteligence jako obor dosáhla v posledních 20 až 30 letech nebyvalého rozmachu.

25 Na poli spekulací o problému strojové mysli uvažovali již filosofové 17. stol., jako např. Descartes, Pascal či Hobbes. Skutečný rozvoj takových úvah ale přinesl až pokrok ve výpočetní technice ve 30. letech 20. stol. Mařík, 1993, s. 15.

26 Tamtéž, s. 18.

27 Tamtéž.

Z českých publikací to můžeme nejlépe vidět na sborníku Umělá inteligence 1-6 (ACADEMIA Praha). Ten v průběhu let 1993 – 2003 problematiku umělé inteligence zevrubně a soustředěně popisoval a zasazoval do českého prostředí. Záběr byl reprezentativně široký od ontologických problémů umělé inteligence přes modely teorie rozhodování a teorie her až po ontologické inženýrství na sémantickém webu. I v době, kdy tato série vycházela, bylo pojetí umělé inteligence jako oboru poněkud neuspořádané a neučesané, dalo by se říci, že přímo chaotické.

Dnes se situace v žádném případě nezměnila a pojednání o umělé inteligenci nebo alespoň o jednom z jejích pojetí v plném rozsahu není možné. Jde o mozaiku filosofických, matematických, lingvistických a inženýrských otázek, poznatků a praktik, které si vytvářejí vlastní paradigmaty a ve kterých je obtížné se zorientovat.

Nebylo by umělé inteligence bez rozvoje počítačových jednotek. Exponenciální nárůst objemů všech druhů pamětí a rychlostí operací umožnil vniknout počítačům do všech oblastí života od průmyslu přes zábavu až po umění a komunikaci. Vizionářských cílů Dartmouthské konference (1956 v Dartmouth College, New Hampshire) pro rok 1970 dosaženo nebylo, ale nadšení pro rozvoj nového oboru bylo všeobecné a přínosy jednoznačné. Předpovídané úspěchy pro rok 1970 byly tyto:

- Počítač odhalí nové významné matematické teorémy.
- Počítač porozumí přirozenému jazyku a bude sloužit jako překladatel.
- Počítač bude schopen komponovat hudbu na úrovni klasiků.²⁸

Tyto požadavky se sice uskutečnit nepodařilo, ale svět se změnil. Práce Alana Turinga, Noberta Wienera, Johna von Neumanna a mnoha matematiků, logiků, inženýrů a průkopníků mnoha jiných oborů stály u zrodu a vývoje nové oblasti lidského poznání – světa virtuality a umělé inteligence. Zaměřme se nyní na Alana Turinga.

28 Mařík, 1993, s. 19.

2. Alan Turing

Alan Turing (23. 6. 1912 – 7. 6. 1954) studoval matematiku, věnoval se ale rovněž filosofii (byl ovlivněn především Fregem, Russellem a Whiteheadem). Od 22 let působil jako asistující profesor na univerzitě v Cambridgi. V roce 1938 narukoval do britské armády a významně se podílel na prolomení šifrovacího kódu německých ponorek. Po ukončení války se Turing s dalšími vědci věnoval vytváření univerzálního počítačového stroje.²⁹³⁰

Ačkoliv Turing během svého života publikoval množství článků, jen jediný byl otištěn ve filosofickém časopise. Je jím *Computing Machinery and Intelligence* z roku 1950.³¹ Zde se Turing jako první zamýšlí nad tím, zda mohou počítače myslet a jak se o této jejich hypotetické schopnosti přesvědčit. Turing se pomocí svého testu, který popisuje ve zmíněné eseji, snažil otázku, zda mohou stroje myslet, převést z oblasti filosofických a lingvistických spekulací na exaktnější úroveň. Tvrdí, že počítač budeme moci prohlásit za myslící tehdy, když jeho chování nebudeme schopni rozeznat od chování člověka.³² Následuje stručný rozbor této významné eseje.

Esej *Computing Machinery and Intelligence*

Když bylo Turingovi pouhých 38 let, napsal jeden z nejvlivnějších filosofických textů 2. pol. 20. stol. – *Computing Machinery and Intelligence*. Hned v úvodu eseje pokládá otázku „*Mohou stroje myslet?*“³³, upozorňuje ale na nejasnost pojmů „stroj“ a „myslet“. O jejich vysvětlení se v tuto chvíli nijak nepokouší a namísto toho nahrazuje původní otázku novou, a sice zda mohou stroje uspět v imitační hře.

Imitační hra je Turingovou koncepcí a není jisté, zda tuto hru někdy někdo zkoušel hrát.³⁴ V eseji jsou představeny dvě verze imitační hry (později se pro ni vžilo označení Turingův test):

29 Tvrdý, 2014, s. 17.

30 Pro podrobný popis života Alana Turinga mohu doporučit bibliografii *Alan Turing: The Enigma* (A. Hodges, 1992) nebo *Turing: A Natural Philosopher* (1997), která je stručným úvodem do jeho myšlení. Poslední zmíněnou bude memoárová kniha *Alan M. Turing* (1959), kterou publikovala jeho matka.

31 Tamtéž.

32 Zároveň je se svým přístupem k inteligenci považován za behavioristu a funkcionalistu.

33 Turing, 1950, s. 433.

34 Tvrdý, 2014, s. 28.

- **Základní verze:** Hru hrají tři lidé – muž (A), žena (B) a tazatel (C), jehož pohlaví není podstatné. Tazatel se nachází v jiné místnosti než účastníci A a B a komunikuje s nimi nejlépe prostřednictvím na stroji psaných vzkazů, jejichž autoři jsou označeni jako X a Y. Úkolem tazatele je zjistit identitu X a Y (tedy buď $X=A$ a $Y=B$, anebo $X=B$ a $Y=A$), přičemž A musí předstírat, že je žena, a B se má snažit tazateli pomáhat nejlépe tím, že mluví pravdu.³⁵
- **Rozšířená verze:** Nyní roli A hraje počítač. Na pohlaví B v tomto případě nezáleží a úkolem C je rozpoznat, který z účastníků je člověk a který stroj. Počítač se samozřejmě snaží co nejvěrněji napodobit člověka (proto např. když je mu zadána matematická úloha, vyčká s odpovědí 30 sekund, ačkoliv výsledek zná takřka ihned). Pokud alespoň ve 30 % odpovědí počítač tazatele přesvědčí, že odpovídá reálný člověk (a nikoliv počítač), pak tento stroj Turingovým testem prošel.^{36,37}

Turing tedy redukuje hypotetickou inteligenci stroje na jeho schopnost napodobit chování člověka, resp. jeho psaný projev. Schopnost uspět v imitační hře pak považuje za důkaz, že tento stroj myslí. Testu se mohou zúčastnit pouze digitální počítače, které Turing charakterizuje následovně:

- Jejich paměť (*store*) se dá donekonečna rozšiřovat.
- Obsahují řídicí jednotku (*control unit*), která vykonává zadané úkoly.
- Na správnou činnost řídicí jednotky podle instrukcí paměti dohlíží část zvaná kontrola (*control*).³⁸

V následující, 5. části eseje se Turing zabývá technickými specifiky digitálních počítačů. Z filosofického hlediska je zajímavá až 6. část, kde Turing přichází s 9 možnými námitkami proti schopnosti strojů myslet a vyvrací je.

35 Turing, 1950, s. 433 – 434.

36 Tamtéž, s. 434.

37 7. 6. 2014 při soutěži na univerzitě v Readingu v Anglii se poprvé v dějinách podařilo „prolomit“ Turingův test. Počítačový program se vydával za 13-ti ročního chlapce z Ukrajiny – Eugena Goostmana. Tento program přesvědčil porotce ve 33 %. Ovšem i tento průlom sebou nese plno „nejasností“ nebo „kritik“. Jednou z nich je ta, že se program vydával za ukrajinského chlapce, který neumí moc dobře anglicky, takže jeho odpovědi byli „krkolomné“ a občas až nesmyslné.

38 Tamtéž, s. 437 – 439.

Turingovy námitky

Aby Turing předešel kritice, ve své eseji podává protiargumenty k 9 možným námitkám proti imitační hře a myšlence inteligentních strojů. Některé z těchto protiargumentů jsou však dosti slabé a proto se zmíním pouze o dvou. Jako první v krátkosti zmíním námitku teologickou, která se zabývá tématem myšlení jakožto lidské schopnosti. Jako druhou, zajímavější a pro tuto práci důležitější, bude námitka „argument z vědomí“ se bude zabývat strojem, jeho schopnostmi a problémy týkající se vědomí stroje.

Teologická námitka

Znění: *„Myšlení je schopností lidské nesmrtelné duše. Bůh dal nesmrtelnou duši každému muži a ženě, nikoliv však živočichovi nebo stroji. Proto žádný živočich ani stroj nemůže myslet,“*.³⁹

Turingovo vyvrácení: Toto tvrzení je v rozporu s doktrínou o Boží všemohoucnosti. Sice existují určité skutečnosti, které Bůh zjevně nemůže změnit (např. matematické pravdy a fyzikální zákony), jistě je ale oprávněn vložit duši, a tedy i schopnost myslet, do libovolného živočicha nebo stroje. Může tak učinit buď přímo, anebo i prostřednictvím člověka-vědce. Turing poznamenává, že jakákoliv argumentace pomocí náboženských dogmat je pro něj nepřijatelná, na námitku náboženského charakteru ale odpovídá ve stejném výkladu.⁴⁰

Argument z vědomí

Znění: *„Dokud stroj nebude schopen napsat sonet nebo zkomponovat koncert na základě vlastních myšlenek a emocí, nikoliv jako výsledek náhodného výběru symbolů, nemůžeme se shodnout na tom, že stroj rovná se mozek – to znamená, stroj musí být schopen dílo nejen napsat, ale rovněž vědět, že ho napsal,“*.⁴¹

39 Tamtéž, s. 443. Překlad autora. Původní znění: „Thinking is a function of man's immortal soul. God has given an immortal soul to every man and woman, but not to any other animal or to machines. Hence no animal or machine can think,“.

40 Tamtéž.

41 Tamtéž, s. 445. Překlad autora. Původní znění: „Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain-that is, not only write it but know that it had written it,“. Tuto definici přednesl na své přednášce G. Jefferson (Jefferson 1949, s. 1110)

Tato námitka obsahuje tři různá tvrzení: stroj si neuvědomuje sám sebe, není schopen pociťovat emoce a není schopen vytvářet umělecká díla.

Z této definice můžeme shrnout 3 základní problémy týkající se vědomí a stroje:

1. Stroj není schopen uvědomit si sám sebe, neboli mít vědomí
2. Žádný stroj nemůže cítit strach, smutek nebo radost či uspokojení ⁴²
3. Stroj není schopen vytvořit umělecké dílo

S bodem číslo tři to v dnešní době není tak úplně pravdivé. Existuje mnoho programů, které dokážou vytvořit či vygenerovat hudbu, literaturu i obrazy.

(Turingovo vyvrácení: Počítač je schopen na dobré úrovni diskutovat o poezii, což značí, že může porozumět umění (to Turing dokládá úryvkem rozhovoru mezi tazatelem a počítačem, v němž společně analyzují Shakespearův Sonet č. 18). Na námitku ohledně neschopnosti stroje uvědomovat si sám sebe Turing odpovídá, že to, jestli má stroj vědomí, můžeme s jistotou říci pouze v případě, že sami budeme tímto strojem. Stejně tak nelze s jistotou určit, zda jsou emoce skutečně prožívané nebo pouze předstírané (to platí nejen u strojů, ale rovněž u lidí).⁴³

Ráda bych zmínila, že přestože v současné době počítače dokážou „neuvěřitelné“ věci, tak nedokážou „mít“ typicky lidské vlastnosti jako být milý, mít smysl pro humor nebo dokonce milovat. Všechny výše jmenované vlastnosti jsou emočního charakteru. Ještě před pár desítkami let bychom do tohoto výčtu mohli zařadit i „neschopnost“ učit se ze zkušenosti, ale v současné době už toto tvrzení neplatí a je překonáno.

Jako příklad mohu uvést používání internetového prohlížeče, kterým je třeba Google. Ten si je schopen zapamatovat jaké odkazy jsem preferovala v předešlém vyhledávání a v budoucnosti tomu přizpůsobuje podobné odkazy. Stejně tak to funguje například s marketingem a reklamami, které jsou přesně a cíleně mířené pro naše potřeby, které jsme již v minulosti vyhledávali.

⁴² Sem můžeme doplnit jakékoliv emoce, které jsme schopni pociťovat

⁴³ Tamtéž, s. 445 – 447.

Kritika eseje

Turingova esej *Computing Machinery and Intelligence* se stala jednou z nejvlivnějších filosofických esejí 2. pol. 20. stol. a setkala se s ohromující odezvou jak pozitivní, tak i negativní. Turingův test a Turingův stroj jsou dodnes rozebíranými tématy mezi odborníky i veřejností. Mnohé komentáře k esejí vykazují známky nepochopení jejího smyslu, což pramení nejspíše z faktu, že v době uveřejnění textu mělo přímou zkušenost s počítači nebo alespoň jasnou představu o nich jen málo jedinců. Uvedme nyní některé z prvních kritik, kterých se Turingův text dočkal.

a) Leonard Pinsky

Jedním z prvních kritiků byl kanadský genetik Leonard Pinsky⁴⁴, který Turingův test parodoval. Navrhl předložit počítači Turingovu esej a pak jej vyzvat k přemýšlení o stroji, který by přemýšlel. Podle Pinského musí počítač nutně selhat. Pinsky využil této příležitosti k útoku na „terapeutické filosofy“, kteří tvrdili, že filosofové dlouhou dobu nemohli v metafyzických otázkách uspět kvůli nesprávnému použití rozumu. Tím, že by počítač v Pinského testu selhal, by ukázal podobnost s těmito neúspěšnými filosofy.⁴⁵

b) Wolfe Mays

Dalším kritikem byl britský filosof Wolfe Mays⁴⁶, který řeší sémantický problém se slovesem „myslet“ a substantivem „stroj“. Místo pojmu „stroj“ navrhuje použít termín „robot“ ve významu sestrojené umělé entity, která by byla schopna myslet, a tedy byla jakýmsi mezičlánkem mezi člověkem a strojem.

Sloveso „myslet“ podle Maysa neznamená „uspět v imitační hře“, ale jde o pojem vyhrazený pouze pro lidi. Mays upozorňuje na to, že Turing se dopouští zásadní chyby používáním termínů psychologického jazyka v souvislosti s neživými bytostmi, pro něž je vyhrazen jazyk fyzikalistický.⁴⁷

44 Mezi jeho největší úspěchy patří např.: výzkum androgenní rezistence, poruch sexuálního vývoje a nozologie malformačních syndromů. Také ovšem napsal „brožuru“ DO MACHINES THINK ABOUT MACHINES THINKING? 1951

45 Tvrď, 2014, s. 70.

46 Mays je známý pro své úsilí o zavedení fenomenologie v Anglii, byl tedy zakladatelem Britské společnosti pro fenomenologii a redaktorem jeho časopisu.

47 Tamtéž, s. 71 – 73.

Samozřejmě nikoho v této době nemohlo napadnout, že počítače se budou do budoucnosti stále jen zmenšovat, nikoliv zvětšovat. Stejně tak, že telefony již nebudou potřebovat žádné kabely a představa dotykových obrazovek by byla v té době opravdu hodně velké science fiction.

c) Keith Gunderson

Keith Gunderson⁴⁸ paroduje Turingův test pomocí své hry, kterou nazývá „šlapaná“. Tato očividně nesmyslná hra vychází z Turingovy imitační hry, jen v ní nejde o komunikaci pomocí psaných vzkazů, nýbrž o šlapaní na nohu. V základní verzi má hráč C za úkol zjistit, zda mu šlape na nohu žena nebo muž, v rozšířené verzi potom, zda stejnou činnost provádí člověk nebo stroj, nazvaný rock-box. Po vzoru Turinga otázku, zda stroj dokáže napodobit člověka, nahrazuje otázkou, zda stroj dokáže uspět ve šlapané.

Gunderson se tak snaží dokázat, že Turingův závěr je chybný. Stroje sice možná dokáží uspět v imitační hře, to ovšem nedokazuje, že by měly schopnost myslet. Úspěch v Turingově testu je možné považovat jen za jeden z projevů inteligence (je to pouze schopnost napodobovat lidské chování v psané konverzaci, inteligence však zahrnuje mnohem více dovedností). Gunderson rovněž upozorňuje na to, že Turing svá tvrzení zakládá pouze na výsledcích činnosti, ve skutečnosti však záleží rovněž na způsobu, jakým se člověk nebo stroj k výsledku dostal – v tom se totiž od sebe obě tyto bytosti zásadně liší.⁴⁹

„Počítač, který prošel T-testem, sice neumí uvařit čaj nebo otevřít láhev piva, ale může poradit jak na to. Imitační hra je tak dostatečným důvodem pro přisouzení ostatních projevů inteligence.“⁵⁰

48 Keith Gunderson je profesorem filozofie na University of Minnesota. Napsal esej *Mentality and Machines*, které bylo poprvé vydáno v roce 1985. Jedná se o obecný esej o filozofii mysli, zaměřený na filozofické a psychologické otázky týkající se skutečných i imaginárních robotů a strojů.

49 Tamtéž, s. 76 – 78.

50 Tvrďý, s. 79

d) Richard Purtil

Purtillova kritika spočívá v tvrzení, že snaha rozlišit odpovědi člověka od odpovědí stroje je marná, neboť ve skutečnosti mají všechny lidský původ. Počítač totiž odpovídá na základě dat, která do něj vložil programátor, zprávy hráče A v imitační hře tedy vlastně nejsou zprávami od počítače, nýbrž od jeho programátora. Purtil nevyvrací možnost, že někdy vědci sestrojí počítače schopné myšlení, tvrdí však, že takový úspěch, vzhledem k současnému stavu poznání, nemůže přijít v blízké době.^{51,52}

e) P. H. Millar

Millar⁵³ vznáší vůči Turingově eseji hned 2 námitky:

1. Turing ve svých úvahách nezohledňuje individualitu lidského účastníka imitační hry. Počítač má sice napodobovat lidskou inteligenci, není ale stanoveno, či intelligence to má být – nikdo přece nemůže zpochybnit tvrzení, že jednotliví lidé se v mentálních schopnostech výrazně liší. Pokud bychom imitační hru nechali hrát počítač a dítě nebo mentálně zaostalého dospělého, je pravděpodobné, že by nad nimi měl počítač navrch.
2. Turing neuvádí postup, jakým lze dosáhnout sestrojení počítače, který by byl schopen myšlení.⁵⁴

f) James H. Moor

J. H. Moore je profesorem intelektuální a morální filosofie. Roku 1985 vytvořil dokument s názvem „Co je to počítačová etika?“ který ho etabloval jako jednoho z průkopnických teoretiků v oblasti počítačové etiky. Také rozsáhle psal o Turingově testu. Jeho výzkum zahrnuje studium filosofie umělé inteligence, filosofie mysli, filosofie vědy a logiky.

51 Tamtéž, s. 80 – 82.

52 R. Purtil byl profesorem filosofie na Západní Washingtonské univerzitě a také autorem fantasy a sci-fi, kritické literatury faktu o stejných žánrech a různých děl o náboženství a filosofii.

53 P. H. Millar je učitelem a univerzitě v Chesteru, kde vyučuje obor sociální a politická filosofie a v roce 1973 napsal článek: „On the point of the imitation game.“

54 Takový postup ale můžeme považovat za oprávněný – vzhledem k tomu, že esej byla uveřejněna ve filosofickém časopisu, by nejspíše postrádalo smysl ji zatěžovat popisem programovacích postupů. Tamtéž, s. 83 – 85.

Zároveň byl až do roku 2010 šéfredaktorem *Minds and Machines*, recenzovaným akademickým časopisem pokrývajícím umělou inteligenci, filozofii a kognitivní vědu.

Moore jako první řekl, že „úspěch v Turingově testu neznamená definitivní odpověď na otázku, zda mohou stroje myslet,“⁵⁵ a není tedy možné Turingův test považovat za definici myšlení. Moor zároveň přikládá svou vlastní definici, která zní: „Myšlení je zpracování informací pomocí metod, které zahrnují rozpoznávání, představivost, vyhodnocování a rozhodování,“.⁵⁶

Předchůdce Turingova testu

Alan Turing jakožto matematik a konstruktér prvních počítačů mohl s přehledem uvažovat o možnostech strojů a jejich hypotetické kompetenci k myšlení. Jakousi obdobu Turingova testu ale nacházíme už v Descartově Rozpravě o metodě: „*Kdyby existovaly stroje, podobající se našim tělům a napodobující naše úkony a bylo by to mravně možné, měli bychom vždy dva vážné důvody, abychom poznali, že proto ještě nejsou skutečnými lidmi. Prvním důvod je, že by nikdy nemohly užívat slov ani jiných znaků, skládající je, jako činíme my, abychom své myšlenky vložili jiným. Neboť lze dobře chápat, že stroj může být udělán tak, aby pronášel slova, ba dokonce, aby pronášel některá ve spojení s tělesnými úkony, souvisejícími s nějakými změnami jeho orgánů: jako například když se ho dotkneme na určitém místě, aby křičel, že ho to bolí, a podobně. Druhým důvodem je to, že nemůže však být udělán tak, aby slova různě sestavoval a takto odpovídal na vše, co se řekne v jeho přítomnosti, jak to i nejtupější lidé mohou činit,*“.⁵⁷ U Descarta jde pochopitelně o čistou spekulaci, námitky proti schopnosti strojů myslet jsou ale podobné těm, které vyslovovali myslitelé v Turingově době.

55 Havel, 2001, s. 34.

56 Moor, 1976, s. 250. Překlad autora. Původní znění: „To think is to process information in ways which involve recognition, imagination, evaluation and decision,“.

57 Descartes, 1992, s. 41.

3. John Rogers Searle

Motto: *Mysl má víc než jen syntax, má i sémantiku.*

(John R. Searle)

John Searle je jedním z nejvýznamnějších amerických filosofů naší doby a je řazen k proudu analytické filosofie (konkrétně k oxfordské škole). Má velmi zásadní podíl na finální podobě teorie mluvnických aktů. Je kritikem a odpůrcem postmoderní filosofie a literární teorie. Jeho významným přínosem je jeho kritika kognitivní vědy. Ta je založena na kritice komputacionismu a na zásluze o znovuoobjevení intencionality v rámci analytické filosofie.

Roku 1984 publikoval paradigmatický článek *Minds, Brains and Programs* Tento článek je brán jako první pokus o odmítnutí Turingova testu.

Toto pojetí tvrdí, že „*správně naprogramovaný počítač disponuje myslí, rozumí a má kognitivní stavy*“, ⁵⁸

Rozeberme nyní první tři, pro tuto práci stěžejní kapitoly Searlovy knihy ***Mysl, mozek a věda***.⁵⁹ V této knize Searle podává námitky vůči Turingově testu a vysvětluje, proč není možné, aby stroje myslely.

Problém mysli a těla

Searle upozorňuje na rozpor, který pozorujeme ve světě – současná věda tvrdí, že celý svět má materiální povahu a je složen z fyzikálních částic, my však vnímáme svou mysl, své emoce, své svobodné rozhodování, zkrátka mentální entity, které se nám jeví jako hmotné. Prostřednictvím své mysli působíme na vnější svět a ten zase působí na naši mysl. Vystává otázka, jak cokoli nehmotného může ovlivňovat hmotné a naopak.

Tradičně bylo toto téma pojímáno jako zkoumání vztahu duše (nebo mysli) a těla, Searle navrhuje jeho přeformulování pro dnešní dobu – zkoumání vztahu mysli a mozku.⁶⁰ Je důležité zmínit, že o fungování lidského mozku toho i v dnešní době víme stále málo, ať už z filosofického nebo neurobiologického hlediska.

⁵⁸ Tvrdý 2014, s. 88

⁵⁹ *Minds, Brains and Science: The 1984 Reith Lectures* (sbírka přednášek; 1984); česky: *Mysl, mozek a věda* (1994)

Mozek řadíme do skupiny nejdůležitějších orgánů lidského těla, ovšem oproti srdci nebo plicím je také jeden z nejsložitějších a „nejtajnějších“ orgánů.

Podle Searla je řešení problému vztahu mysli a mozku snadné, mnoha lidem ale činí problém na něj přijít, protože o problémech dnešní doby se hovoří v termínech filosofie 17. století. Současný vědec si tudíž musí vybrat, zda bude dualista či monista, což vede ke striktnímu rozlišování mezi fyzickým a duševním světem. Jeden z těchto fenoménů je pak potlačen ve prospěch druhého, v současnosti je situace taková, že mentální entity jsou podceňovány a důraz je kladen na materiální skutečnost. Searle fakt, že mnoho vědců popírá existence mentálních fenoménů, vysvětluje tím, že tyto fenomény mají čtyři vlastnosti, které se zdají neslučitelnými s objektivní vědou. Jsou to vědomí, intencionalita, subjektivita a mentální kauzace. Searle je považuje za reálné vlastnosti psychického života člověka. Rozebereme postupně každou z nich. Nejprve však musíme přijmout Searlovo tvrzení, že procesy, které se odehrávají v mozku na úrovni nervových buněk, způsobují mentální stavy, a tyto stavy se rovněž v systému buněk realizují.⁶¹

Vědomí

Vědomí je základním rysem lidského života, umožňuje totiž jazyk, myšlení, emoce apod. Díky vědomí vnímáme okolní svět, svoje postavení v něm a také vše, čemu právě věnujeme svoji pozornost. Zdá se však nepochopitelné, že vědomí, které vnímáme jako nemateriální, může být vlastností fyzikálních objektů.

Searle ale tvrdí, že existenci vědomí v tkáních můžeme vysvětlit na základě biologických procesů. Stejně jako život je i vědomí výsledkem elektrochemické aktivity. Teprve poté, co pochopíme funkci nervových buněk, jejich vzájemný vztah a aktivity, které mezi sebou mají, tak pochopíme, jak vědomí funguje a jak je „možné“.⁶²

a) Intencionalita

Slovo intencionalita pochází z latinského slova intentio, tedy úsilí, zaměření. Intencionalitou Searle rozumí schopnost našich mentálních stavů vztahovat se k vnějšímu světu. Podobně jako v předchozím případě vyvstává otázka, jak mohou fyzikální částice v mozku reprezentovat cokoliv vnějšího.

60 Searle, 1994, s. 14.

61 Tamtéž, s. 23.

62 Tamtéž, s. 16, 24 – 25.

Searle existenci intencionality vysvětluje pomocí biologie. Biologické procesy způsobují mentální stavy a tyto stavy se realizují rovněž v biologických systémech.⁶³

b) Subjektivita

Mentální stavy jsou nutně subjektivní, je totiž vyloučeno, aby někdo jiný např. cítil bolest způsobenou tím, že jsme se uhodili do hlavy. Mentální stavy může zažívat vždy jen ten, v jehož mozku se odehrávají, jiným lidem je může leda zprostředkovat pomocí jazyka. Moderní věda ale klade na veškeré své zkoumání požadavek objektivity, tedy požaduje, aby její předměty zkoumání byly přístupné všem a aby je všichni vnímali stejně. Searle poznamenává, že věda by neměla ze skutečnosti vylučovat subjektivitu, její existence totiž může být považována za objektivní fakt. Věda by měla uznat, že součástí objektivního světa jsou rovněž osobní mentální procesy odehrávající se v mozcích lidí a některých živočichů.⁶⁴

c) Mentální kauzace

Naše mentální stavy prostřednictvím našeho chování a jednání působí na fyzický svět a pozměňují ho. A zároveň to, co se děje ve vnějším světě, působí na naše city a ovlivňuje tedy naši mysl. Filozofové ale nejsou schopni vysvětlit, jak mentální může ovlivňovat fyzické a naopak.

Rovněž kauzaci lze podle Searla vysvětlit prostřednictvím biologie. Každý mentální stav, jakožto vlastnost mozku, můžeme popsat na dvou rovinách – psychologické a fyziologické. Každou myšlenku můžeme pojmut jako biochemickou reakci mezi nervovými buňkami. Na vnější svět tedy nepůsobí žádné nehmotné fenomény, nýbrž náš nervový systém.⁶⁵

Závěr

Celou kapitolu Searle shrnuje tvrzením, že mysl a tělo se sice vzájemně ovlivňují, nejedná se ale o dvě odlišné skutečnosti. Mysl je vlastností mozku, který se součástí těla. Toto pojetí spojuje dva na první pohled neslučitelné názory – fyzikalismus a mentalismus.⁶⁶

63 Tamtéž, s. 16 – 17, 25 – 26.

64 Tamtéž, s. s. 17, 26 – 27.

65 Tamtéž, s. 17, 27 – 28.

66 Tamtéž, s. 28.

Mohou počítače myslet?

Ve druhé kapitole se Searle zapojuje do diskuse o potenciální schopnosti strojů myslet, kterou započal Turing. Proti jeho názorům se také Searl vymezuje. Tvrdí, že ačkoliv je fungování mysli a její vztah s tělem snadno vysvětlitelné pomocí biologie (jak jsme viděli v předchozí kapitole), současní odborníci z různých vědeckých odvětví přirovnávají činnost mozku k činnosti počítače a mysl pak srovnávají s počítačovým programem. Lidská mysl by podle tohoto pojetí byla napodobitelná, jen se prozatím nepodařilo sestavit vhodný počítačový program, který by umožnil libovolnému počítači chovat se, jako by měl mysl. Počítač vybavený tímto programem by měl potom stejně mentální a intelektuální schopnosti jako člověk. Intelligence je tak redukována na schopnost fyzicky zacházet se symboly. Přístup, který chápe mysl jako počítačový program a mozek jako stroj, který tento program spouští, Searle nazývá silnou umělou inteligencí.⁶⁷

Searle uvádí názory některých zastánců silné umělé intelligence na potenciální schopnost strojů myslet. Prvním z nich je vědec Herbert Simon, který říká, že už byly vynalezeny stroje, které dokáží myslet. Jsou jimi digitální počítače, které mají dle Simona myšlenky stejně jako lidé. Druhým vědcem, kterého Searle uvádí, je výzkumný pracovník v oblasti informatiky a kognitivní psychologie Alan Newell. Ten tvrdí, že intelligence je výhradně záležitostí fyzikálních operací se symboly. Jako posledního Searle jmenuje vědce Marvinu Minského, který je přesvědčen o tom, že „*příští generace počítačů bude už tak inteligentní, že budeme vůbec rádi, když si nás budou vydržovat alespoň jako domácí zvířata*“.⁶⁸

Zatímco námitky Searlových předchůdců vůči možnosti, že by počítače mohly myslet, jsou často postaveny na upozornění na nedokonalost soudobých počítačů, Searle útočí na samotnou definici digitálního stroje. Proto jeho argument není nijak závislý na úrovni technického pokroku a nemůže být ani v budoucnosti vyvrácen.

Searlova námitka proti možnosti, aby stroje myslely, se dá shrnout do prostého: „*Mysl má víc než jen syntax, má i sémantiku*“.⁶⁹ Searle vychází z běžně přijímané definice digitálního stroje, která říká, že takový stroj provádí určité operace pomocí abstraktních symbolů. Počítač ale není schopen tyto symboly vnímat jako reprezentující nějaký obsah, pro stroj se zkrátka symbol nepojí s žádným významem.

67 Tamtéž, s. 29 – 30.

68 Tamtéž, s. 30 – 31.

69 Tamtéž, s. 33.

Naproti tomu pro lidskou mysl má každý symbol nějaký obsah – příkladem může být fakt, že když zaslechneme nebo někde vidíme napsané nějaké slovo, automaticky si představíme předmět, který toto slovo označuje. Lidská mysl tedy pracuje s fenomény, které mají určitou formu či syntax (toho je schopen i počítač), přiřazuje jim ale i obsah či sémantiku (to počítač už z definice činit nemůže).⁷⁰

Kognitivní věda

Kognitivismus je vědecká disciplína, která si dala za cíl vysvětlit pochody v mysli a mozku jinak než prostřednictvím fyziologie nebo psychologie. Dříve se o to pokoušely směry jako behaviorismus, teorie her, kybernetika, teorie informace nebo strukturalismus.⁷¹

Kognitivní věda vychází z analogie člověka a počítače a čerpá z poznatků kognitivní psychologie a umělé inteligence. Snaží se o popis lidského mozku a jeho činnosti jako o popis systému, který zpracovává nějaké informace. Nejde zde tedy ani o zkoumání fyziologických procesů, ani mentálních stavů a pochodů.

Searle s kognitivismem nesouhlasí, uznává však, že je pro mnoho lidí lákavý.

Uvádí pro to hned 4 důvody:

1. Člověk má sklony mozek přirovnávat k aktuálně nejdokonalejšímu technickému vynálezu. Tak se dříve činnost mozku popisovala mj. jako činnost katapultu, hodinového stroje, telegrafu nebo hydraulického stroje. Dnes za nejmodernější techniku považujeme počítače a roboty, a proto mozek vysvětlujeme jako počítač.⁷²
2. Důvěryhodnost kognitivních věd je podporována dvěma typy psychických evidencí. Za prvé, řešení intelektuálních úkolů trvá různým lidem různě dlouho, stejně tak existuje rozdíl v čase v případě člověka a počítače, což by mohlo znamenat, že mozek i počítač pracují na podobných principech. Za druhé, člověk se při používání jazyka řídí formálními pravidly stejně jako počítač.⁷³

70 Tamtéž, s. 32 – 33.

71 Searle, 1994, s. 45 – 46.

72 Tamtéž, s. 47 – 48.

73 Tamtéž, s. 48 – 49.

3. Předpokládá se, že mentálním výkonům předchází teorie, a tedy v lidském mozku musí být nějaký „program“, který tyto výkony umožňuje.⁷⁴
4. Ke kognitivismu tíhnou lidé, kteří nejsou schopni vztahu mysli a mozku porozumět jiným způsobem.⁷⁵

Argumenty proti kognitivismu

John Searle se snaží kognitivismus vyvrátit. Uvádí proto, v čem se lidé svým chováním liší od fungování počítače.

Jedna ze základních tezí kognitivní vědy říká, že člověk se řídí podle pravidel stejně jako počítač. Počítač se ale pravidly řídí ryze formálně a nerozumí jejich významu. Zároveň výsledek jeho činnosti napovídá, že se počítač s jistotou řídí podle pravidla, které k tomuto výsledku vede. Člověk ale může výsledku dosáhnout i v případě, že se žádným pravidlem neřídí, anebo se řídí jiným pravidlem, než se předpokládá. V případě počítače má ono pravidlo podobu několika za sebou jdoucích kroků, které když stroj zopakuje, dosáhne výsledku. U člověka nejde o přesné znění pravidla, nýbrž o jeho obsah, význam. A samozřejmě díky přítomnosti svobodné vůle člověk může pravidlo porušit.⁷⁶

Podle Searla tedy existují dva způsoby chápání sousloví „řídít se podle pravidel“ - doslovný a metaforický. V případě počítačů se jedná pouze o metaforu, neboť jejich činnost sice zvenjšku vypadá stejně jako chování člověka, ve skutečnosti však toto chování pouze napodobují a obsahu pravidel nerozumějí. Stejně je to i s používáním jazyka – počítač se „řídí“ syntaktickými pravidly jen formálně, rozhodně jinak než člověk.

Kognitivisté rovněž tvrdí, že člověk při myšlení zpracovává informace a totéž dělá i počítač. I zde vidí Searle dvojí význam. Informace zpracovávané člověkem jsou nutně součástí jeho mentálních procesů, a to buď vědomě, anebo nevědomě. Počítač však žádné mentální procesy nemá, pouze je napodobuje. První způsob nazývá Searle „psychickým zpracováním informací“, druhý „jakoby zpracování informací“.

74 Tamtéž, s. 49.

75 Tamtéž.

76 Tamtéž, 49 – 51.

Upozorňuje na nebezpečí záměny těchto dvou pojetí, čehož se kognitivismus očividně dopouští.⁷⁷

Dále se Searle snaží vyvrátit tezi, že každému smysluplnému chování předchází teorie skrytá uvnitř subjektu. Upozorňuje na fakt, že mnohé mentální výkony lze vysvětlit fyziologicky, a tudíž zde žádné výpočty či pravidla nejsou potřeba, děláme je podvědomě a zcela automaticky i bez jakékoliv teorie. Kdyby však tyto úkony měl dělat počítač, musel by se na teoretický základ (určitý soubor pravidel) spolehnout.⁷⁸

Argument čínského pokoje

Podstatu své námitky Searle čtenářům přibližuje pomocí myšlenkového experimentu s místností, v níž je umístěno několik košů s čínskými znaky a podrobný manuál pro práci s nimi psaný v jazyce, kterému účastník myšlenkového experimentu dobře rozumí (tento manuál však popisuje pouze to, jakým způsobem se mají symboly řadit za sebe, jak se na který znak odpovídá apod., význam těchto znaků nevykládá). Aktér myšlenkového experimentu (který neumí čínsky) se má tímto manuálem řídit, a když jsou mu do místnosti dopraveny další čínské znaky, má na oplátku z místnosti vysílat ven znaky podle přiloženého návodu. Tento aktér tedy vlastně vykonává funkci počítačového programu, do jehož databáze bylo uloženo určité množství čínských znaků a pravidla pro formální zacházení s nimi. Tomuto programu jsou zadávány otázky a on na ně odpovídá. Aktér, pokud následuje manuál, může na otázky pokládané v čínštině odpovídat přesně a gramaticky správně v tomtéž jazyce. Rozhodně ale nemůžeme říci, že nyní tento člověk umí čínsky – nikdy mu totiž nebyl osvětlen význam znaků, které mechanicky používal.⁷⁹

Z vnějšího pohledu se tedy zdá, že počítač, do nějž vložíme dobře propracovaný program pro čínštinu, tento jazyk ovládá, tedy že je dostatečně inteligentní na to, aby jím mohl mluvit. Ve skutečnosti ale počítač pouze postupuje podle manuálu, aniž by rozuměl tomu, co dělá. Člověk, který mluví čínsky, musí znát význam jednotlivých znaků a slov (sémantiku), počítač je schopen pouze formálních operací se symboly (syntaxe).⁸⁰

77 Tamtéž, s. 52 – 54.

78 Tamtéž, s. 54 – 57.

79 Tamtéž, s. 33 – 34.

80 Tamtéž.

Searle uvádí námitky, které různí filosofové a vědci vznesli proti jeho argumentu čínského pokoje. Jednou z těchto námitek je, že počítač sice nerozumí jednotlivým znakům, rozumí ale čínštině jako celku. To by ale znamenalo, že počítač samovolně odvodil ze syntaxe sémantiku, což není možné (stejně jako člověk nemůže porozumět významu čínských symbolů jen na základě pravidel pro jejich třídění).⁸¹

Další námitka upozorňuje na možnost vložit program pro čínštinu do robota, který by se pohyboval světem. Pokud by byl schopen komunikovat s lidmi mluvícími čínsky na úrovni rodilého mluvčího, autoři této teorie by to považovali za důkaz, že robot ovládá čínštinu. Searle opět namítá, že robot nemůže přejít od syntaxe k sémantice, takže to, že umí čínsky, je pouze zdání.⁸²

Digitální počítače a myšlení

Po uvedení argumentu čínského pokoje se Searle vrací k otázce, zda stroje mohou myslet. Jestliže stroj chápeme jako „*fyzikální systém schopný provádět jisté operace*“⁸³, pak do této definice spadá i člověk, a tedy existuje stroj, který myslí. Původní otázka se však netýkala lidí, a proto je vhodné ji přeformulovat takto: může nějaký stroj, který byl stvořen člověkem, myslet? Kdybychom dokázali sestrojit stroj, který by byl svou stavbou shodný s člověkem, pak by odpověď zněla ano. Jde tu ale o sestrojení digitálního počítače, nikoliv umělého člověka.

Digitální počítač je definován jako stroj, který realizuje nějaký počítačový program. V tomto smyslu můžeme za digitální počítač považovat i člověka. Na lidském myšlení se ale projevuje schopnost rozumět významům symbolů, kterou stroj nemá. Myšlení je tedy něco víc než pouhé uskutečňování počítačového programu, a proto stroj, který dokáže pouze toto uskutečňování, nemůže nikdy myslet.⁸⁴

81 Tamtéž, s. 36.

82 Tamtéž.

83 Tamtéž, s. 37.

84 Tamtéž, s. 37 – 39.

Závěr

V závěru kapitoly Searle prostřednictvím jednoduchých výroků shrnuje nejdůležitější teze této části a vyvozuje z nich čtyři závěry:

P1: Mentální stavy jsou způsobovány mozkiem.

P2: Z ryze formálního nelze vyvozovat žádný obsah.

P3: Počítačové programy se z definice vyznačují formální strukturou.

P4: Mysl má mentální obsahy, tedy nemá pouze formální strukturu, ale rovněž obsahovou.

Z1: Počítač nemůže myslet, a to ani v případě, že mu dáme sebedokonalejší program.

Z2: Počítačový program nemůže zapříčinit, aby mozek způsoboval mysl.

Z3: Mysl může být způsobována jedinež kauzálními silami, které jsou přinejmenším srovnatelné s kauzálními silami mozku.

Z4: Hypotetický stroj s mentálními stavy by musel být vybaven něčím více než jen počítačovým programem.⁸⁵

Předchůdci Searlova argumentu

Podobný myšlenkový experiment jako Searle provádí už někteří novověcí filosofové, např. Leibniz. Ten se v 17. oddílu své *Monadologie* zamýšlí nad povahou lidského vnímání. Jeho argumentace je velmi podobná té Searlově a oba myslitelé zaujímají k možnosti existence myslícího stroje skeptický postoj.

Rovněž matematik Hartley Rogers předjal Searlův argument čínského pokoje tím, že se zabýval metaforickým popisem Turingova stroje, a to celých 20 let před samotným Searlem. Rogersův argument zahrnuje muže, který je uzavřen v krabici, kde má pouze stůl, tužku a papír. Samotná krabice na sobě má ještě otvory pro vstup a výstup. Jeho úkolem je přiřadit k číslům, která dostane na vstupu, jiná čísla, která odevzdává u výstupu.⁸⁶

⁸⁵ Tamtéž, s. 40 – 43.

⁸⁶ Rogers, 1959, s. 115 – 117.

Alternativa kognitivismu

Searle kognitivismus odmítá, nabízí ale jeho vhodnou alternativu. Navrhuje, abychom si mysl a mentální procesy začali představovat nikoliv jako fungování počítače, nýbrž jako biologické procesy.

Při jejich popisu zjistíme, že mají dvě roviny – neurofyzilogickou a intencionální. Žádná další rovina zde není třeba – ani ta, kterou nabízí kognitivismus.⁸⁷

I navzdory Searlově kritice je v současné době kognitivismus stále zastáván, jeho podoba se však změnila. V diskusích o problematice mysli a těla v rámci kognitivní vědy se dnes významně uplatňuje přístup vtělené kognice (*embodied cognition*) či vtělené mysli (*embodied mind*). Tento přístup odmítá striktní oddělování mysli a těla (a někdy též ztotožňování mysli s mozkem), tvrdí totiž, že myšlení probíhá uvnitř těla, které je umístěno v nějakém prostředí, a že i samotné myšlení se vztahuje k věcem, které jsou vůči němu vnější – fyzické tělo i okolní prostředí tedy myšlení významně ovlivňují, myšlení se neodehrává ve vzduchoprázdnu. Nelze ho zkoumat, aniž bychom zároveň nebrali v potaz tělo a prostředí, v nichž k výkonům mysli dochází. Podobně v případě umělé inteligence nejde čistě jen o to, umět vytvořit správný počítačový program, ten totiž rovněž nemůže být nezávislý na svém okolí.⁸⁸

Turing a Searle – shrnutí

Alan Turing a John Searle představují dva pohledy na umělou inteligenci a její možnosti – zatímco Turing, který se konstrukcí počítačů sám zabýval, je přesvědčen o tom, že je v lidských silách zkonstruovat takový stroj, který by nejen napodoboval lidské myšlení, ale přímo sám myslel, Searle je k tomuto názoru kritický. Upozorňuje na nepatřičnost přirovnávání lidského myšlení k fungování stroje, které je základní myšlenkou kognitivismu. Searle argumentuje tím, že myšlení je biologický proces sestávající z různých mentálních procesů, které jsou způsobeny mozkem a rovněž se v mozku odehrávají – počítač mozek nemá, alespoň ne ten skutečný, biologický, který je tohoto schopen. S problematikou mysli a těla se Searle vypořádává tvrzením, že na veškeré mentální procesy a obsahy lze pohlížet dvěma způsoby – vědecky jako na

87 Tamtéž, s. 57 – 60.

88 Russel, Norvig; 2016, s. 1026.

biochemické procesy a interakci mezi neurony, a rovněž ze subjektivního pohledu jako na jakési nehmotné skutečnosti. Jedná se však o dvě strany téže mince.

Turing redukuje schopnost myslet na schopnost uspět v jeho imitační hře, tedy vlastně na schopnost věrohodně (ale nikoliv naprosto přesvědčivě, když vezmeme v potaz požadovanou úspěšnost 30 %) napodobovat lidskou psanou komunikaci. To je podle Searla příliš málo. I kdybychom připustili, že je možné veškeré lidské myšlení takto redukovat, to, že počítač něco věrně napodobuje, ještě neznamená, že tomu rozumí – to se Searle snaží dokázat pomocí tzv. argumentu čínského pokoje. Dochází k závěru, že počítač ve své činnosti sice vykazuje formální znaky myšlení, jak ho pojímá kognitivismus (tj. řídí se podle pravidel a zpracovává informace), avšak není schopen porozumět obsahu dané činnosti. A protože pro myšlení jsou klíčové obě složky, není možné tvrdit, že počítač skutečně myslí. Zde je ale možno uvést námitku sestavenou na základě Turingova pokusu o vyvrácení argumentu z vědomí, a sice že je velmi obtížné dokázat, že počítač obsahům skutečně nerozumí, dokud sami nebudeme tímto počítačem.

V současnosti asi nejrozšířenější přehledová publikace na téma umělé inteligence, *Artificial Intelligence: A Modern Approach* od Russela a Norviga, uvádí celkem 8 definic umělé inteligence, seřazených do 4 kategorií podle toho, co podle nich musí stroje umět napodobit. Turingův přístup směřuje k požadavku napodobení lidského chování, kognitivní věda zase požaduje od umělé inteligence schopnost myslet jako člověk. Kromě toho se však objevují ještě další dva přístupy kladoucí důraz na racionalitu, a to buď na racionální myšlení, anebo racionální jednání.⁸⁹ Je tedy patrné, že názory na to, co činí počítač inteligentním, nejsou ani dnes zdaleka jednoznačné. Situaci nenapomáhá ani fakt, že dosud nebyla ustanovena všeobecně uznávaná definice lidské inteligence, z níž se by se případně dalo odvodit vymezení inteligence umělé.

Můžeme ale říci, že ze všech 4 základních kategorií, uvedených výše, se v současnosti nejvíce rozvíjejí přístupy, které definují umělou inteligenci na základě schopnosti racionálního jednání. Myslicí počítač je podle nich racionálním aktérem – je to stroj, který jedná vždy tak, aby dosáhl nejlepšího možného výsledku. Důraz na racionalitu je podle mnohých vědců výhodnější než snaha napodobit člověka, neboť

89 Russel, Norvig; 2016, s. 2.

člověk je ve svém myšlení a jednání ovlivněn bezpočtem faktorů, jako jsou třeba emoce, morálka, náboženství apod., a ne vždy jedná racionálně, předvídatelně a tím nejlepším možným způsobem.

Naproti tomu naprogramovat počítači racionalitu je poměrně neproblematické a přináší to kýžené výsledky. To, že je vhodnější soustředit se spíše na racionální jednání než na racionální myšlení, bývá obhajováno poukázáním na fakt, že nejlepšího možného výsledku je možné dosáhnout i jinak než pomocí myšlení, a že někdy je myšlení spíše na obtíž.⁹⁰

90 Russel, Norvig; 2016, s. 4 – 5.

4. Kybernetický funkcionalismus versus biologický naturalismus

V současné době, která je značně ovlivněna kybernetickou revolucí v pol. 20. st., je kognitivismus velmi rozšířenou teorií. Svůj základ nachází ve funkcionalismu, konkrétně v kybernetickém funkcionalismu. Zastáván je ale i přístup zcela odlišný od funkcionalistických koncepcí – biologický naturalismus. Z rozdílnosti obou přístupů často vznikají rozsáhlé diskuse, které končí vzájemným nepochopením a neschopností dobrat se jakéhokoliv závěru. Proto se nyní pokusíme mezi sebou kybernetický funkcionalismus a biologický naturalismus porovnat a kromě rozdílů snad nalézt i nějaké podobnosti.

Obě teorie se zabývají zkoumáním rozdílů mezi přírodními, přirozenými strukturami a strukturami umělými, lidmi vytvořenými. Zatímco biologický naturalismus zdůrazňuje jedinečnost mozku a myšlení a tyto fenomény vysvětluje na základě fyziologie, podle kybernetického funkcionalismu je každý předmět vysvětlitelný jakožto stroj, automat, jehož činnost je možné při dostatečné technické zdatnosti věrně napodobit. Podle této koncepce tedy lze úspěšně simulovat i lidské myšlení, protože mozek se nijak neliší od člověkem vyrobených počítačů.⁹¹

Zastánci biologického naturalismu (a to jak z řad biologů, tak i filosofů) popírají, že by bylo možné mentální procesy živých bytostí přirovnávat k činnosti počítačů, natož je napodobovat. Upozorňují na fakt, že sklon vysvětlovat vztah mozku a mysli jako vztah hardwaru a softwaru je historicky podmíněný a úzce souvisí s moderním technokratickým přístupem. Zároveň stavba biologických struktur je od počítačů velmi odlišná, a proto není v žádném případě možné je srovnávat. Funkcionalisté tuto argumentaci považují za nevědeckou, naivní a romantizující. Podle nich se jedná pouze o antropocentrickou snahu dodat lidskému mozku a mysli punc jedinečnosti.

Podívejme se nyní podrobněji na jednotlivé přístupy a nabídneme jejich kritiku.

91 Bartoš, 2013, s. 110.

Kybernetický funkcionalismus

Funkcionalismus je teorie, která vzniká v 2. pol. 20. stol. a stojí na přesvědčení, že jakýkoliv předmět či živá bytost nejsou definovány svým vzhledem a fyzikálními vlastnostmi, nýbrž funkcí, již vykonávají.

Z klasické podoby funkcionalismu se odvozuje kybernetický funkcionalismus, jehož hlavní pilíře jsou stavěny na analogii mezi stroji a myslícími bytostmi. Každý (i biologický) systém lze chápat jako konečný automat, který provádí nějaké výpočty na základě určitých algoritmů.

Automat může být vyroben z jakéhokoliv materiálu, nezáleží ani na jeho velikosti a tvaru, důležitá je jediná funkce – a ta je pro všechny počítače stejná, protože se žádné dva automaty neliší v tom, co umějí a co nikoliv. Daniel Hillis to nazývá principem univerzálního počítače. A jelikož jsou všechny automaty stejné (ve své funkci), je možné je mezi sebou libovolně nahrazovat, stačí mít jen dobře sestavený program, který je pak možné realizovat na libovolném počítači. Tato myšlenka se označuje jako ontologická univerzalita funkcionalismu. Z ní vyplývá, že lidské myšlení lze provádět na jakémkoliv konečném automatu, pokud mu dodáme správný program.⁹²

Další základní tezí kybernetického funkcionalismu je princip vícere realizace, který vyplývá z ontologické univerzality a spočívá v substrátové neutrálnosti funkcí. Libovolný proces lze provádět na jakémkoliv konečném automatu, neboť jeho fyzikální vlastnosti jsou pro správnou činnost irelevantní. Tedy i činnosti, které běžně vykonávají biologické struktury, lze provádět na syntetickém, nebiologickém substrátu.

Vysoká míra abstrakce a zmíněné principy spolu s rozšířenou metaforou lidského mozku jakožto počítače vedly k tomu, že funkcionalisté prakticky přestali rozlišovat mezi původní strukturou a její pouhou simulací. To ale není to jediné, co by bylo možné funkcionalismu vytknout. Ačkoliv se prezentuje jako čistě materialistická a monistická teorie, pracuje s dvojicí software-hardware, která nápadně připomíná karteziánskou dvojici duše-tělo. Jestliže je možné lidskou mysl „přenést“ do libovolného počítače, znamenalo by to, že je od těla oddělitelná, což by rovněž odpovídalo dualistickému pojetí. Princip vícere realizace navíc může být chápán jako aktualizace myšlenky stěhování duší.⁹³

92 Tamtéž, s. 112 – 113.

93 Tamtéž, s. 114.

Kritika principu vícere realizace

Již Daniel Dennett si byl vědom toho, že princip vícere realizace se může s absolutní platností realizovat jedinečně v případě umělých entit. Biologické „počítače“ jsou vždy vázány na svůj substrát, jinak řečeno mentální procesy jedince určitého živočišného druhu se mohou odehrávat výlučně v těle příslušníka tohoto druhu. Je to způsobeno tím, že řídicí systémy vznikající v průběhu evoluce vždy určitým způsobem navazovaly na systémy jim předcházející, vycházely z nich a rozvíjely je. Bez nich by tyto dokonalejší systémy nemohly fungovat.

Je možné napodobit činnost specializované části takového systému (např. v případě protetických náhrad částí těla), nikoliv však tento systém jako celek.⁹⁴

Proti principu vícere realizace se staví i teorie „scaling laws“⁹⁵, kterou jako první vyslovil už Galileo Galilei. Ten tvrdí, že vzhledem k tomu, že hmotnost je druhou mocninou lineární velikosti nějakého předmětu či tvora a síla třetí mocninou velikosti, není možné stejný předmět nebo bytost řádově zvětšovat či zmenšovat. Aby totiž mohly s novými rozměry správně fungovat, bylo by nutné pozměnit jejich stavbu nebo je vytvořit z jiného materiálu.⁹⁶

Pokud se tedy budeme pokoušet funkce určitého konečného automatu přenést do jiného, který se od něj zásadně liší, může se stát, že se tato funkce změní – tedy že nový automat bude fungovat jinak než ten původní. Každá funkce má v tomto ohledu určité meze, za nimiž už buďto není sama sebou, nebo není možné ji identifikovat. Dennett tento princip ukazuje na myšlenkovém experimentu – kdybychom se dostali do kontaktu s mimozemskými bytostmi, jejichž mysl by pracovala milionkrát rychleji než naše, tyto bytosti by naše schopnosti porovnávaly se svými a jen stěží by nás považovaly za inteligentní. Stejně jako my považujeme za nemyslicí stromy – je však možné, že jejich myšlení je jen natolik pomalé, že nejsme schopni ho zpozorovat. Z toho plyne, že funkce myšlení se musí odehrávat v určitém časovém rozmezí, tedy že funkce (přinejmenším funkce myšlení) není převoditelná na libovolný konečný automat.⁹⁷

94 Tamtéž, s. 114 – 116.

95 „Scaling laws“ neboli měřítkové zákony, jsou vztahy mezi fyzickými veličinami, ve kterých se všechny fyzické veličiny objevují ve smyslu mocností. Mocnina x je vyjádřena ve tvaru x^α , kde α je reálné číslo.

96 Tamtéž, s. 76 – 77.

97 Dennett, 2004, s. 62 – 63.

Když to shrneme, princip vícere realizace platí pouze za předpokladu, že funkce, která je přenášena do jiného počítače, probíhá v čase a prostoru srovnatelném s funkcí, kterou tento substituující počítač původně vykonával, a že oba počítače umožňují funkci zhruba stejnou sít' vztahů.

Rovněž se zdá problematickou teze, že myslící bytosti by mohly být sestrojeny z nebiologických materiálů. Na základě Dennettovy věže (kap. 1.2.1) totiž můžeme tvrdit, že bytosti stojící ve vyšších patrech věže (v našem případě úplně nejvýše, jelikož se má jednat o tvory s myslí – tedy gregoryovské tvory či lépe řečeno jejich napodobeniny) mohou existovat pouze díky těm, které jsou postaveny níže.

Pokud bychom tedy chtěli sestrojit myslící bytost, museli bychom zrovna tak vytvořit simulaci všech tří typů tvorů jí předcházejících. Je ovšem otázkou, zda je možné autonomní darwinovské tvory, kteří jsou schopni se sami množit, vybudovat z nebiologických materiálů. Podle chemika Richarda Smalleyho, laureáta Nobelovy ceny, to možné není.⁹⁸

Biologický naturalismus

Podstatou biologického naturalismu je přesvědčení, že biologické entity se svou stavbou výrazně liší od entit uměle vytvořených. Přírodní bytosti sice zpracovávají informace, avšak jinak než digitální počítače, a proto není možné lidský mozek přirovnávat k počítači. Tento názor zastával i Alan Turing, tvrdil totiž, že i malá odchylka ve velikosti nervového vzruchu může způsobit velký rozdíl ve velikosti impulsu vycházející z neuronu, což v počítači nastat nemůže.⁹⁹

Digitální stroje mohou mít naprogramovány určité stavy, které se podobají těm, které nacházíme u živých bytostí. Počet těchto stavů může být v závislosti na pokroku techniky nesmírně vysoký, ale vždy je omezený, tyto stavy jsou jasně definované a zrovna tak možné způsoby, jak jeden stav přechází v druhý.

98 Bartoš, 2013, s. 118 – 119.

99 Turing, 1950, s. 456.

Naproti tomu biologičtí tvorové v průběhu evoluce nabyly dvou odlišných, avšak na sobě závislých typů procesů – modulárních, které se odehrávají stejně přesně jako procesy počítačové, a hérakleitovských,¹⁰⁰ které způsobují jisté odchylky. To proto, že je zde vždy velké množství aktérů, které vzájemně pojí nahodilé vazby. Díky tomu nelze stavy biologických entit přesně definovat ani předpovídat, je jich nekonečné množství a jeden stav ve druhý může přecházet bezpočtem různých způsobů. Můžeme tedy říci, že biologické entity jsou zároveň digitálními stroji i stroji analogovými (kontinuálními).¹⁰¹

Kritika biologického naturalismu

Tvrzení, že biologické „stroje“ pracují digitálně i analogově, může být ovšem chybné. Ve skutečnosti mohou být veškeré jejich stavy přesně definované a bez odchylek, jen my, lidé, nejsme (zatím) schopni jim porozumět a vidět v nich řád. Protože jsou tyto procesy nesmírně složité a vyžadují abstrakci (když totiž zkoumáme myšlení, zkoumáme neurony, a sice neurony jako celky, přičemž pomíjíme fakt, že jsou tvořeny bezpočtem dílčích komplexních částí), máme sklony je pojímat jako nepředvídatelné a odlišné od procesů odehrávajících se v počítači. Ve skutečnosti se však možná v tomto ohledu lidský mozek od počítače zas tolik neliší.¹⁰²

Ačkoliv se lidská mysl jako celek může zdát býti zcela odlišnou od digitálních počítačů, můžeme její strukturu rozkládat na stále drobnější a základnější stavební jednotky, které vždy pracují jako stroj. Lze tedy tvrdit, že lidský mozek je počítač skládající se z milionů hierarchicky uspořádaných počítačů.¹⁰³

100 Používáme zde rozlišení dr. Bartoše, jež představil ve své dizertační práci. Modulární procesy můžeme přirovnat k činnosti digitálních strojů, jsou pravidelné, mají jasnou strukturu a hierarchii. Naopak herakleitovské procesy připomínají procesy analogových strojů a vyznačují se jakousi nahodilostí, chaotičností, nepředvídatelností – mezi tyto procesy lze zařadit např. Přelévání tekutiny nebo plápolání ohně. Bartoš, 2013, s. 23 – 24.

101 Bartoš, 2013, s. 120.

102 Tamtéž, s. 121.

103 Tamtéž, s. 122.

Kybernetický funkcionalismus a biologický naturalismus – shrnutí

Kybernetický funkcionalismus i biologický naturalismus jsou protikladné vědecké a filosofické přístupy ke vztahu počítačů a lidské mysli. Funkcionalismus se snaží lidský mozek popsat pouze na základě jeho funkcí a odhlíží od jeho morfologie. Pro funkcionalisty nezáleží na tom, zda je nějaký systém utvořen z organického materiálu či vyroben z materiálu syntetického, v obou případech je možné tento systém převést na konečný automat. Jelikož jsou tedy všechny systémy ve své funkci shodné, lze jakýkoliv proces provádět na libovolném konečném automatu. To by znamenalo, že lidské myšlení je realizovatelné v uměle vyrobeném počítači, stačí mít správný počítačový program. Pokud však vezmeme v úvahu teorii scaling laws a princip vývoje organismů podle Dennettovy věže, můžeme možnost libovolné záměny procesů na počítačích prohásit za nepravděpodobnou.

S principem víceré realizace rovněž zásadně nesouhlasí biologický naturalismus. Ten přikládá velký význam substrátu, na němž procesy probíhají, a odmítá možnost proces probíhající na organickém substrátu realizovat na substrátu syntetickém, ba dokonce i na organickém substrátu odlišného biologického druhu. Podle naturalistů je počítač schopen reprezentovat určitý omezený počet stavů, mezi nimiž přechází rovněž omezeným počtem způsobů. Tyto stavy a přechody mezi nimi jsou jasně definované a přehledné. Avšak u živých bytostí může dojít k nekonečnému množství stavů a bezpočtu způsobů přechodu, stavy jsou nejasné, špatně definovatelné, nepředvídatelné. Toto přesvědčení však může pramenit z nedostatečného vědeckého poznání, je možné, že ve skutečnosti mají veškeré stavy živých organismů povahu modulárních procesů.

Na těchto teoriích se jen těžko hledá něco společného. Můžeme však pozorovat, že jak funkcionalismus, tak i biologický naturalismus chápou činnost počítače a lidské myšlení jako určité procesy, a ačkoliv naturalisté odmítají možnost biologické procesy realizovat na syntetickém substrátu, jejich pojetí těchto procesů se podobá chápání digitálního a analogového počítače. Z toho můžeme vyvodit, že i biologický naturalismus připouští alespoň vzdálenou podobnost mozku s automatem, ačkoliv tvrdí opak.

5. Závěr

V této bakalářské práci jsem se snažila rozebrat některé filosofické aspekty umělé inteligence. Tato problematika je nicméně tak rozsáhlá a stále se rozvíjející, že nejsem schopna v tomto rozsahu podat její ucelený obraz, snad ale může posloužit alespoň jako základní vhled do ní. Věnovala jsem se rozlišením lidské mysli od zvířecí a od uměle vytvořené, a to z různých filosofických stanovisek.

Nejprve jsem ve zkratce nastínila problém vztahu duše (mysli) a těla, který se poprvé objevuje v antické řecké filosofii, do středu zájmu filosofů se dostává v novověku (např. u Pascala, Descarta, Leibnize) a dále je rozvíjen v osvícenství (např. de La Mettrie). Od 20. st. je pod vlivem rychlého vědeckého pokroku pojmán jako vztah mysli a mozku. V této době také otázka, zda může člověk sestrojít myslící stroj, kterou se filosofové zabývali již od novověku, přestává být pouhou spekulací, neboť skutečně dochází ke konstrukci strojů na vysoké úrovni, které jsou schopny zpracovávat informace podobně jako člověk.

K úvahám o potenciální schopnosti počítačů a robotů myslet je nezbytné definovat, co vlastně rozumíme pod pojmy mysl, myšlení, inteligence. Proto jsem hned první kapitolu věnovala této problematice. Zjistili jsme, že definovat tyto pojmy je nesmírně obtížné, uvedla jsme ale alespoň některé z možných definic. Lidskou inteligenci jsem porovnávala s inteligencí živočichů a umělých entit. Z této kapitoly bych vyzdvihla vysvětlení tzv. Dennettovy věže, metafory pro čtyři vývojové stupně organismů směřující k rozvoji intencionality.

Druhá a třetí kapitola pojednávají o postoji Alana Turinga a Johna Searla k možnosti existence myslících umělých entit. Turing, který se sám věnoval vytváření počítačů, ve své eseji *Computing Machinery and Intelligence* rozvíjí myšlenku, že v budoucnosti lidé budou schopni vytvářet počítače schopné myslet. Za kritérium pro posouzení, zda daný počítač myslí, navrhuje úspěch tohoto počítače v imitační hře – tedy stroj můžeme považovat za myslící, jestliže nejsme schopni z jeho odpovědí rozpoznat, zda se jedná o stroj či člověka.

Naproti tomu Searle je vůči potenciálním myslícím strojům skeptický. Zastává pozici tzv. slabé umělé inteligence, která počítače považuje za užitečnou pomůcku pro zkoumání lidské mysli, možnost, že by počítače skutečně myslely, ale odmítá. Ve své knize *Mysl, mozek a věda* vyvrací myšlenku, že je možné sestrojít umělé entity schopné myšlení jako lidé, a podává námitky vůči Turingově imitační hře.

Searle tvrdí, že zatímco lidská mysl zcela přirozeně pracuje s obsahy (významy), počítače toho nejsou schopny, pouze formálně napodobují lidské mentální procesy. To je ukázáno na myšlenkovém experimentu čínského pokoje. Dále se Searle domnívá, že je možné veškeré mentální stavy a procesy vysvětlit jako biologické (přesněji neurofyziologické) jevy, a tudíž není nutné v úvahách rozlišovat mysl (duši) a mozek (tělo), neboť mysl je způsobena mozkem a také se v něm odehrává.

V práci nebyl opomenut ani moderní kognitivistický přístup ke zkoumání potenciálních myslí umělých entit. Ten popisuje počítač jako analogii lidskému mozku, protože se oba při své činnosti řídí podle pravidel a zpracovává informace. Searle podrobuje obě tvrzení kritice a dochází k závěru, že v obou případech jsou ve hře dva různé významy – doslovný a metaforický. Počítač se řídí podle pravidel jen metaforicky, protože nerozumí jejich obsahům, a taktéž metaforicky zpracovává informace, neboť zpracovávané informace nemohou být součástí jeho mentálních procesů jako v případě člověka.

Ve čtvrté kapitole jsme pojednali o rozdílných přístupech kybernetického funkcionalismu a biologického naturalismu. Funkcionalismus definuje entity na základě jejich funkce, přičemž ta je u každého digitálního počítače (i lidského mozku) stejná. Na základě principů ontologické univerzality funkcionalismu a vícere realizace zastánci tohoto směru tvrdí, že lidské myšlení lze provádět na jakémkoliv digitálním stroji z libovolného materiálu. Naproti tomu biologický naturalismus jakékoliv přirovnávání mozku k počítači odmítá a naopak zdůrazňuje jedinečnost lidského myšlení a jiných mentálních procesů. Vzhledem k naprosté odlišnosti obou směrů v diskusi mezi jejich zástupci není možné dojít shody, takové diskuse se nesou v duchu nepochopení.

Pozornost byla zaměřena především na typy otázek, které v souvislosti s umělou inteligencí vyvstávají, a metod zkoumání lidské mysli, které je k rozvoji umělé inteligence nezbytné.

Z různých teoretických základů jsme tedy prozkoumali otázky, zda mohou počítače myslet, jaký je vztah mysli a mozku, čím se liší přírodní (přirozené) od umělého, jak rozpoznat inteligenci u lidí, zvířat a umělých entit apod. Naopak jsme se nevěnovali futurologickým vizím umělé inteligence – toto téma by vydalo na samostatnou práci.

Na závěr se ještě zmiňme o Průmyslu 4.0 neboli o 4. průmyslové revoluci, která bude využívat kyberneticko - fyzikální systémy. Tyto řízené robotické systémy údajně nahradí lidskou sílu. V některých oborech, kde je pro člověka práce velmi těžká a ničí mu zdraví, to bude jistě přínos. Ale čím více těchto systémů bude v provozu, tím více lidí bude bez práce. Samozřejmě, pro lidi budou k dispozici jiné pracovní pozice, bude potřeba mnoho inženýrů a počítačových specialistů. Lidé již nebudou muset vykonávat pracovní úkony fyzicky namáhavé nebo jednoduché, s opakující rutinní pracovní činností (např. ve zpracovatelském průmyslu). Dále budou ohroženy pracovní profese, jako jsou úředníci, pokladníci, prodavači ve všech odvětvích, obsluhy jednoduchých strojů nebo dělnické profese.

Ráda bych práci uzavřela svým vlastním názorem, který tkví v tom, že superintelligence UI jednou určitě překoná tu lidskou.¹⁰⁴ Ovšem otázkou je, co nás potom čeká? Nahradí nás v práci a my budeme nezaměstnaní? Budeme ji moci nebo budeme vůbec ještě schopni ji kontrolovat? A dovolí nám to? Bude na naší straně nebo proti nám? Možná nás jako druh už nebude potřebovat a protože bude mít moc nás zničit, tak to udělá. Vždyť my se chováme stejně třeba ke zvířatům, která můžeme vyhubit, tak jak jsme to již některým druhům udělali. Superintelligence třeba nebude mít žádné morální zábrany všechny ostatní druhy zničit, narozdíl od lidí. Konec konců dokud budou stroje programovat lidé, tak jim také mohou některé své špatné vlastnosti vložit do programu.

Ze všech pěti kategorií otázek, které si filosofové i vědci kladou v souvislosti s umělou inteligencí, považuji za nejzávažnější ty etické. Autonomní rozhodování je problém, s kterým se budeme setkávat, až místo nás budou rozhodovat stroje. Dnes má odpovědnost za rozhodování člověk a za své špatné rozhodnutí nese následky. Jaké následky ale ponese stroj, který udělá špatné rozhodnutí a způsobí třeba smrt člověka? Jakou morálku tyto stroje budou mít a budou mít vůbec nějakou? To vše jsou velmi zapeklité otázky, na které zatím neznáme odpověď.

104 Již nyní je vyvíjen typ silné umělé inteligence Seed AI, která bude schopna se sama zdokonalovat.

Seznam použité literatury:

HAVEL, Ivan M. Úvod. In: MAŘÍK, Vladimír a kolektiv. *Umělá inteligence (1)*. Praha: Academia, 1993. ISBN 80-200-0496-3.

HAVEL, Ivan M. Přirozené a umělé myšlení jako filozofický problém. In: MAŘÍK, Vladimír a kolektiv. *Umělá inteligence (3)*. Praha: Academia, 2001. ISBN 80-200-0472-6.

HAVEL, Ivan M. Úvod. In: MAŘÍK, Vladimír a kolektiv. *Umělá inteligence (6)*. Praha: Academia, 2013. ISBN 978-80-200-2267-9.

DENNETT, Daniel Clement. *Druhy myslí: k pochopení vědomí*. Praha: Academia, 2004. Mistři vědy. ISBN 80-200-1177-3.

SEARLE, John. Minds, brains, and programs. *The Behavioral and Brain Sciences*, 1980, 3(3), pp. 417-457

SEARLE, John R. *Mysl, mozek a věda*. Přeložil Marek NEKULA. Praha: Mladá fronta, 1994. ISBN 80-204-0509-7.

TVRDÝ, Filip. *Turingův test: filozofické aspekty umělé inteligence*. Praha: Togga, 2014. Scholia (Togga). ISBN 978-80-7476-043-3

MOORE, G. E., 1925. A Defence of Common Sense. In: J. H. MUIRHEAD, ed. *Contemporary British Philosophy*. London: Allen and Unwin, s. 194

HUME, David. *Zkoumání o lidském rozumu*. Přel. Josef MOURAL. Praha: Svoboda, 1996. Filozofické dědictví. ISBN 80-205-0521-0.

SEARLE, John R., 1992. *The Rediscovery of the Mind*. Cambridge: MIT Press.

PUTNAM, Hilary, 1988a. Much Ado about Not Very Much. *Daedalus*. Roč. 117, č.1

HOBBS, Thomas, 1988. O tělese. In: Vojtěch BALÍK (přel.) *Výbor z díla*. Praha: Svoboda

DESCARTES, René. *Rozprava o metodě*. Přel. Věra SZATHMÁRYOVÁ-VLČKOVÁ. Praha: Svoboda, 1992. ISBN 80-205-0216-5.

TURING, Alan, 1950. *Computing Machinery and Intelligence*. *Mind*. roč. 59, č.236, s. 433-460

MOOR, James H., 1976. *An Analysis of the Turing Test*. *Philosophical Studies*. roč. 30, č.4, s.2 49 – 257.

SEARLE, John R., 1980 b. *Minds, Brains and Programs*. *Behavioral and Brain Sciences*. Roč. 3, č. 3, s. 417 – 424.

MORWITZ, Vicki. *Consumers' Purchase Intentions and their Behavior*. Now Publishers, 2014. ISBN: 9781601988805

MAYS, W., 1952. *Can Machines Think?* *Philosophy*. Roč. 27, č. 101, s. 154.

SEARLE, John R., 1994b. Searle, John R. In: Samuel Guttenplan, ed. *A Companion to the Philosophy of Mind*. Oxford: Blackwell, Rogers, Hartley, 1959. *The Present Theory of Turning Machine Computability*. *Journal of the Society for Industrial and Applied Mathematics*. Roč. 7, č. 1, s. 115 – 117.

POLÁK, Michal. *Filosofie mysli*. Praha: Triton, 2013. ISBN 978-80-7387-742-2.

KASSIN, Saul M. *Psychologie*. Brno: Computer Press, 2007. ISBN 978-80-251-1716-3.

RUSSELL, Stuart J. a Peter NORVIG. *Artificial intelligence: a modern approach*. 3rd ed. Harlow: Pearson Education, c2016. ISBN 978-1-29215-396-4

ROGGERS, Carl R. *A Theory of Therapy, Personality, and Interpersonal Relationships: As Developed in the Client-centered Framework*. McGraw-Hill, 1959

SIMMONS, Aaron, J. *The New Phenomenology: A Philosophical Introduction*. Bloomsbury Academic; 1st edition, 2013.

Seznam použitých webových stránek:

<https://plato.stanford.edu/entries/chinese-room/>

<https://www.iep.utm.edu/chineser/>

<https://citaty.net/citaty-o-inteligenci/>

<https://vesmir.cz/cz/on-line-clanky/2015/06/intelligence-co-vlastne-je.html>

https://is.muni.cz/th/yrgxq/disertacka_komplet_verze_xyz.pdf

<https://www.zive.cz/deepmind/sc-702/default.aspx>

<https://www.csfd.cz/film/283747-kod-enigmy/prehled/>

http://casopis.mensa.cz/veda/intelligence_a_jeji_mereni.html

<https://web.archive.org/web/20071210043312/http://members.aol.com/NeoNoetics/MindsBrainsPrograms.html>