

Západočeská univerzita v Plzni

Fakulta filozofická

Bakalářská práce

Filosofické aspekty umělé inteligence

Petr Štas

Plzeň 2017

Západočeská univerzita v Plzni

Fakulta filozofická

Katedra filozofie

Studijní program Filozofie

Studijní obor Filozofie

Bakalářská práce

Filosofické aspekty umělé inteligence

Petr Štas

Vedoucí práce:

Mgr. Michal Polák, Ph.D.

Katedra filozofie

Fakulta filozofická Západočeské univerzity v Plzni

Plzeň 2017

Prohlašuji, že jsem práci zpracoval samostatně a použil jen uvedených pramenů a literatury.

Plzeň, duben 2017

Poděkování:

Velice rád bych poděkoval svému vedoucímu bakalářské práce panu Mgr. Michalovi Polákovi, Ph.D., za odborné vedení práce, toleranci a cenné rady, které mi během psaní této práce poskytl.

Obsah

Úvod	6
1. Problémy s identifikací mysli	8
2. Rozdíl mezi přirozeným a umělým	10
3. Turingův stroj a funkcionalismus Hilaryho Putnama.....	13
4. Imitační hra.....	15
5. Námitky vůči imitační hře.....	16
5.1 Teologická námitka.....	16
5.2 Námitka „hlavy v písku“	17
5.3 Matematická námitka	18
5.4 Argument z vědomí	19
5.5 Argument z různých neschopností.....	20
5.6 Námitka lady Lovelace.....	21
5.7 Argument ze spojitosti nervové soustavy	23
5.8 Argument z neformálnosti chování.....	23
5.9 Argument z mimosmyslového vnímání	23
6. Loebnerova cena	24
7. Silná a slabá umělá inteligence	28
8. Searlův argument čínského pokoje	30
9. Eugene Goostman	32
Závěr	33
Použitá literatura.....	34
Summary.....	35

Úvod

Příroda některé živé organismy obdařila vlastností, která jim dává mimořádné postavení. Touto nadřazenou vlastností je inteligence. V dnešní době inteligence umožňuje některým živým organismům působivě reagovat na všemožné složité události a efektivně je využívat ve svůj prospěch, k pokoření svých cílů. Už od doby, kdy docházelo k rozvoji techniky, si lidé kladli otázku, zdali je možné, aby uměle vytvořené systémy dosahovaly stejných reakcí a chování, jako živé organismy a zdali bychom je mohli na základě toho považovat za inteligentní. Lidé se však těmito otázkami zabývali dokonce i mnohem dříve než byly vynalezeny elektronické počítače. Tyto otázky řešili již v 17. století významní filosofové jako jsou Descartes, Hobbes, či Pascal.¹

Například Descartes v *Rozpravě o metodě* představuje, jak je na základě své nauky o dvojí substanci k myslícím strojům značně skeptický: „A tu jsem se zvláště zastavil u důkazu, že kdyby existovaly takové stroje, jež by měly orgány a vnější vzhled opice nebo jiného nerozumného zvířete, měli bychom důvod se domnívat, že by byly ve všem stejné povahy jako tato zvířata; kdežto kdyby existovaly stroje, podobající se našim tělům a napodobující naše úkony potud, pokud by to mravně bylo možné, měli bychom vždy dva velice vážné důvody, abychom poznali, že proto ještě nejsou skutečnými lidmi. První důvod je, že by nikdy nemohly užívat slov ani jiných znaků, skládající je jako činíme my, abychom své myšlenky vyložili jiným. Neboť lze dobře chápat, že stroj může býti 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 se zeptal, co mu chceme říci, když na jiném místě, aby křičel, že ho to bolí, a podobně; 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.“²

Závěry filozofů 17. století měli samozřejmě pouze filozofickou podobu a zaměřovali se spíše na zodpovězení otázky, zdali mohou stroje myslet, nikoliv na sestavení návodu, jak dosáhnout strojového myšlení. Teprve až ve 30. letech 20. století se některé významné výsledky v matematické logice a v teorii algoritmů staly skutečnými katalyzátory úsilí napodobit intelektuální chování člověka. Velmi značným přínosem byl nástup a velice rychlý rozvoj výpočetní techniky po druhé světové válce. Byly navrhovány a experimentálně ověřovány různé metody, postupy a algoritmy, které umožňovaly napodobovat určité hledisko inteligentního chování. Všechny tyto postupy či algoritmy, které svým způsobem vedou k určitému napodobení projevů inteligentního chování člověka, jsou předmětem zkoumání vědní disciplíny, která je nazývána: „umělá inteligence“.³

I. M. Havel představuje čtyři motivační zdroje, proč bychom se měli zabývat výzkumem umělé inteligence. První je motivace aplikační nebo inženýrská, tato motivace je podle Havla

¹ HAVEL, Ivan M. Úvod. In: MAŘÍK, Vladimír a kolektiv. *Umělá inteligence (1)*. Praha: Academia, 1993. str. 15.

² DESCARTES, René. *Rozprava o metodě*. Praha: Svoboda, 1992. str. 41.

³ HAVEL, 1993. str. 15.

v současné době převládající. Druhá motivace je badatelská, ve které je cílem dovědět se co nejvíce o povaze mentálních procesů. Další je motivace matematická, ve které jde především o čistě teoretické studium logických a matematických systémů, odvozených z metod umělé inteligence. A poslední motivací je přirozená lidská hravost, ve které nás může zajímat, co všechno dovedeme sestrojít a naprogramovat.⁴

Vzhledem k tomu, že tématem mé práce jsou „Filosofické aspekty umělé inteligence“, tak je zřejmé, že mé motivace jsou čistě badatelské. Cílem mé práce je představit základní problémy a pojmy umělé inteligence. V této práci bude nejprve představeno, jaké jsou problémy s identifikací mysli, a osvětlíme si rozdíl mezi umělým a přirozeným. Převážná část práce se bude zabývat prací Alana Turinga, kde budou zmíněny pojmy „Turingův stroj“, „funkcionalismus“ a hlavně „Turingův test“. Pozornost budeme věnovat i Loebnerově ceně, která si dodnes prochází značnou kritikou. Následně bude představen „argument čínského pokoje“ od Johna Searla a jeho rozlišení „slabé a silné umělé inteligence“. Mým dalším cílem, který bude pojednán v závěrečné části práce, je zjistit, zdali existuje software, který by byl schopný zcela projít Turingovým testem ve všech případech s úspěšností 100%.

⁴ HAVEL, Ivan M. Přirozené a umělé myšlení jako filozofický problém. In: MAŘÍK, Vladimír a kolektiv. *Umělá inteligence (3)*. Praha: Academia, 2001. str. 18.

1. Problémy s identifikací mysli

Filip Tvrdý v úvodu své knihy *Turingův Test: Filozofické aspekty umělé inteligence* předkládá tvrzení, podle kterého jsme obklopeni mnoha myslícími, inteligentními entitami. Téměř každý z nás se dokonce s těmito entitami, u kterých zvládneme identifikovat mysl a které považujeme za vědomé, setkává každý den. Z toho vyplývá otázka: Kdo nebo co patří do této kategorie entit, u kterých předpokládáme, že mají mysl?⁵

Do této kategorie patří především ostatní lidé. Není však nutné začínat s žádnými komplikovanými filozofickými úvahami, abychom byli schopni přiznat mysl téměř všem lidem. Tvrdý upozorňuje na problematickou podskupinu, do níž patří někteří mentálně postižení lidé nebo lidé v kómatu a pravděpodobně i děti do určité fáze prenatalního či postnatalního vývoje. Ostatním lidem mysl přiznává a předkládá několik filozofických důvodů, jak je to možné.⁶

Jako první řešení používá argumentaci z pozic zdravého rozumu, která navazuje na nauku o zdravém rozumu (common sense), která se dá nalézt u osvícenského filosofa Thomase Reida nebo analytického filosofa G. E. Moora. Tato nauka tvrdí, že existují tvrzení, o kterých je naprosto zbytečné diskutovat či snažit se je dokázat nebo vyvrátit. Moore je nazval „truismy“ a patří mezi ně rozsáhlá škála tvrzení. Jako příklad Tvrdý uvádí: mám tělo, narodil jsem se a žiji na Zemi, existují dějiny, existují jiní lidé, mám zkušenosti různých druhů, jiní lidé mají zkušenosti různých druhů a podobně. Na závěr uvádí, že každá venkovanka bez vzdělání či malé dítě vědí, že ostatní lidé myslí, a proto může být toto tvrzení považováno za truismus.⁷

Tvrdý poukazuje na další řešení, které je považováno za pragmatické. Toto východisko upozorňuje, že je neúčinné, abychom pochybovali o tom, že mají ostatní lidé mysl. Podle Tvrdeho se pravděpodobně poprvé objevuje u Huma a to v jeho díle *Zkoumání o lidském rozumu*.⁸ Hume tvrdí, že z přehnaného skepticismu: „... nikdy nemůže pocházet žádné trvalejší dobro, pokud setrvává v plné síle a sveřeposti.“⁹

Posledním a nejjednodušším řešením je podle Tvrdeho prosté poukázání na neurální podobnosti mezi jednotlivými lidmi: „Jsme si podobní, vytvoření ze stejného materiálu, máme stejně uspořádané vnitřní orgány a stejné chemické složení. Jsme vybaveni stejnými mozky se stejným nebo obdobným rozložením mozkových center. Vypadáme podobně zvenku i zevnitř, funkčně si odpovídají naše zažívací i oběhová ústrojí, bylo by tedy velmi nepravděpodobné – řekněme nemožné – že bychom se závažně lišili mentálně. Já myslím a nutně předpokládám, že můj podobně zkonstruovaný kamarád myslí též.“¹⁰

⁵ TVRDÝ, Filip. *Turingův test: filozofické aspekty umělé inteligence*. Praha: Togga, 2014. str. 7.

⁶ Tamtéž, str. 7-8.

⁷ Tamtéž, str. 8.

⁸ Tamtéž.

⁹ HUME, David. *Zkoumání o lidském rozumu*. Praha: Svoboda, 1996. str. 217.

¹⁰ TVRDÝ, 2014. str. 9.

Můžeme si povšimnout, že s identifikací myslí u ostatních lidí nemáme značně velké problémy. Avšak komplikace se objeví poté, co se podle Tvrdého pokusíme připsat status myslících bytostí za hranicemi našeho biologického druhu. Všechna tři řešení, která jsme použili na případu jiných lidských bytostí, zde selhávají. Někdo může jednotlivá tvrzení považovat za samozřejmá a někdo jiný je naopak bude vidět jako neobhajitelná. Někteří považují za samozřejmé, že zvířata myslí a že i možná rozumí lidské řeči, na druhé straně je představa myslících zvířat pro některé lidi naprosto absurdní.¹¹

Tvrdý upozorňuje na to, že naše manipulace s pojmy, které různým entitám přisuzují mentální stavy, je značně nepřehledná a nespravedlivá. Důvodem je, že někdy dochází k tomu, že máme přílišnou tendenci aplikovat pojem „mysl“ i na lidi, kteří na to očividně nemají právo. Příkladem jsou lidé v kómatu. Někdy dochází k pravému opaku, kdy tento pojem odebíráme zvířatům, která na něj zřejmě právo mají. Situace se enormně zkomplikuje, pokud se do kategorie entit majících mysl pokusíme zařadit ty entity, u kterých je jejich nárok relativně nedávný, například počítače, stroje a roboty.¹²

Může být stroj schopen myslet stejně jako člověk? Může přirozeně myslet? Nejprve však bude potřeba vysvětlit rozdíl mezi pojmy „přirozené“ a „umělé“. Tímto problémem se budeme zabývat v následující kapitole.

¹¹ TVRDÝ, 2014. str. 9-10.

¹² Tamtéž, str. 11-12.

2. Rozdíl mezi přirozeným a umělým

Snadnou odpovědí na otázku o vztahu přirozeného a umělého, by podle Jana Romportla bylo: „umělé je to, co tu dříve nebylo, člověk si usmyslel, že to vytvoří, a nyní to tu je. Přirozené je pak vše ostatní. Umělé dle této definice jsou tedy kupříkladu hráz, dům, kladivo, počítač, ale třeba i píseň. Přirozené jsou naproti tomu květiny, stromy, hory či lidé.“¹³

Romportl přiznává, že jednoduchost výše uvedené definice začne být problematická, když se začneme zabývat jednotlivými případy. Můžeme považovat za umělý i vzrostlý les, pokud ho před mnoha generacemi člověk vsadil do krajiny vlastníma rukama? A i kdyby měl být považován za umělý, tak je samozřejmě jeho umělost zásadně odlišná od umělosti automobilu. Nejproblematictější je podle Romportla případ samotného člověka, kterého by podle všeho mělo být možno považovat za přirozeného, avšak v mnoha případech existuje okolnost, kdy způsob jeho „vzniku“ naprosto přesně odpovídá definici umělého.¹⁴

Tento rozpor může být dobře překonán, pokud použijeme tři podmínky, které nám nabízí I. M. Havel, k tomu abychom o nějaké věci, vlastnosti či činnosti mohli tvrdit, že je umělá. První podmínkou je: „Existuje nějaká *přirozená* věc, logicky připouštějící duplikaci (v našem případě je to například lidské myšlení, vnímání, rozhodování apod.).“¹⁵ Druhá podmínka: „Existuje *záměr* člověka (nebo týmu) vytvořit duplikát oné přirozené věci.“¹⁶ A třetí podmínkou je: „Došlo k *provedení záměru*, čili proběhl intencionální proces vedoucí od záměru k jeho *realizaci*.“¹⁷

Romportl se domnívá, že první podmínku můžeme vynechat: „Jelikož se snažíme dopátrat nejen toho, co je umělé, ale i co je přirozené, tak předpoklad přirozeného vzoru v charakterizaci umělého nám zavádí definici kruhem.“¹⁸ Následně uvádí, že dokonce mnoho umělých věcí ani nemusí být napodobeninami žádných přirozených vzorů. A tak, i když kladivo může být v určité funkci pouhým napodobením kamene nebo dům napodobením pravěké jeskyně, tak počítač bude těžko napodobováním něčeho přirozeného.¹⁹

Velmi důležitým momentem je zde podle Romportla intencionalita záměru i tvůrčího procesu. Například samotné zrození člověka může být svázáno s intencionálním záměrem jeho početí, ale jeho vlastní proces vývinu se stává dějem, který je zcela samovolný. Stejně tak to platí i u čerstvě vysázených stromků, které jsou vysázeny důsledkem uvědomělého

¹³ ROMPORTL, Jan. Přirozenost umělé inteligence. In: BENEDIKTOVÁ, Marie et al. *Člověk v nových světech*. Plzeň: Vydavatelství Západočeské univerzity v Plzni, 2012. str. 115.

¹⁴ Tamtéž, str. 116.

¹⁵ HAVEL, 2001. str. 31.

¹⁶ Tamtéž.

¹⁷ Tamtéž.

¹⁸ ROMPORTL, 2012. str. 116.

¹⁹ Tamtéž.

záměru a činnosti člověka, proto je možné stromky v tomto stádiu chápat jako umělé, avšak to co z nich v budoucnu učiní vzrostlý les, pochází naprosto odjinud než z lidského konání.²⁰

Následně Romportl představuje myšlenkový experiment s golemem: Představme si, že pokroky ve vědě a technologii nabízejí takové možnosti, díky nimž lze buňku po buňce sestavit člověka (například jako „kopii“ nějakého „přirozeného vzoru“). Ačkoli by toto sestavování prováděl nejspíše nějaký velice složitý stroj navržený člověkem, bylo by to stejné, jako kdyby tento stroj prováděl sestavování jiného složitějšího stroje (například kopii sebe sama)? Byla by vzniklá kopie člověka sama člověkem? A to člověkem umělým, nebo přirozeným?²¹

Je možné říci, že to co je přirozené, pochází z přírody. A to, co samo pochází z přírody, je její součástí. Pak tedy, když lidská mysl, lidská řeč a lidská tvořivost pocházejí z přírody, tak i jejich plody jsou přirozené. Dalo by se říci, že podle těchto tvrzení je golem přirozený? Tato tvrzení však podle Romportla vedou k dalšímu extrému, kdy lze těžko považovat igelitový obal či dálnici za přirozené, alespoň ne potud, pokud chceme za přirozený považovat svůj úsudek.²²

Co se tedy dá nazvat vznikem či zrozením golema z Romportlova myšlenkového experimentu? Tím je právě vnější intencionální a racionální činnost jeho tvůrce, proto snad můžeme golema podle Romportla nazvat člověkem umělým. Podívejme se však na definici slova *fysis*, tedy přirozenosti: „Fysis je vše, co prochází rozením a umíráním, vznikem a zánikem.“²³ Z toho tedy podle Romportla vyplývá, že jakmile golem povstane, tak se stává člověkem, a to právě člověkem přirozeným, protože se *fysis* začne projevovat skrze jeho tělesné ustrojení. Golem, tak bude vždy pro svého tvůrce umělým, avšak sám golem bude ve své *fysis* přirozeným člověkem.²⁴

Jako přirozené by Romportl nazval to, co se vzpírá být uchopeno jazykem. Nebo, jak následně podotýká: „Přirozené je to, u čehož cítíme napětí mezi tím, co jsme chtěli jazykem uchopit, a co jsme jím nakonec uchopili.“²⁵ Ve všech případech, kde se jazyk nestřetává s podstatou toho, co popisuje, můžeme vidět přirozeně jsoucí. Jazykovým uchopením, zde Romportl nemyslí pouhé arbitrární přiřazení slova k věci, jako například u slov „kořen“ či „korálový útes“ ke kořeni a korálovému útesu, ale myslí tím právě možnost představení oné věci v celém svém rozsahu nějaké osobě, která se s ní dosud nesetkala, a to jinak, než jejím pouhým ukázáním. Což se podle něj u kořene rostliny či u korálového útesu pravděpodobně nedá.²⁶

²⁰ ROMPORTL, 2012. str. 116.

²¹ Tamtéž, str. 117.

²² Tamtéž.

²³ KRATOCHVÍL, Zdeněk. *Filosofie živé přírody*. Praha: Herrmann a synové, 1994. str. 6.

²⁴ ROMPORTL, 2012. str. 117.

²⁵ Tamtéž, str. 118.

²⁶ Tamtéž.

Žádné slovní vyjádření nepopíše, jak existuje, ani jak přesně vypadá obyčejné ptačí hnízdo, oblak na obloze či vlna na vodní hladině. Stejně tak ani sebelepší matematický model není schopen něco takového popsat. Romportl přiznává, že je to možné snad pomocí jazyka básnického, ale ten v tomto případě ponechává stranou. Podle jeho názoru nám jazyk racionality i matematický model popisují „podstatu úplně jiných objektů, než jsou ty, s nimiž se přirozeně ve světě setkáváme.“²⁷ Z toho vyvozuje, že přirozené jsou ty jevy, od kterých musíme abstrahovat, pokud chceme slovy vyjádřit jejich podstatu.²⁸

Naopak u věcí a jevů umělých je charakterem jejich soulad podstaty. Tímto souladem je podle Romportla to, co je pomocí jazykového uchopení činí tím, čím jsou a jakým způsobem jsou. U umělých věcí se nesetkáváme s tím, aby se vzpíraly tomu být uchopeny jazykem, právě naopak se k tomu přímo vybízejí, proto se tak s nimi dá mnohem snadněji manipulovat než s věcmi přirozenými. Jazyková deskriptivní abstrakce umělé věci je vlastně totožná s její podstatou, z čehož by se podle Romportla dalo usoudit, že umělé se stává samou abstrakcí sebe sama. Následně uvádí, že pokud budeme pozorně vnímat to, co všechno můžeme na nějakém objektu jazykem uchopit, tak máme často i možnost nakonec poznat, co z přirozeného se nachází v základech mnoha věcí umělých. Jako příklad uvádí dřevěnou desku tesaného stolu, která může být verbalizována možná jen popisem procesu, kterým byla vytesána a také do některých rozměrů své geometrické idealizace. To čím je však dřevo samotné, není možné dále popisovat jinak než tím, že ho přirovnáme k něčemu jinému, nebo pomocí vědecky motivované redukce na výčet jiných objektů, z nichž se dřevo skládá.²⁹

Podle Romportla je rysem přirozené věci to, že jediná její věrná reprezentace je právě tato věc sama o sobě, protože neexistuje žádný zobecňující předpis, podle kterého by bylo možné zestručnit popis oblaku či květiny a přitom by je zachoval v jejich původní podobě a také v jejich původním vztahu ke světu. Na druhé straně vše, co je důležité na nějakém stroji a tvoří jeho podstatu, jsme zcela schopni shrnout dostatečně přesným konstrukčním plánem a schématem řízení.³⁰

Ze všeho výše uvedeného vyplývá, že tu máme dvojí chápání přirozeného a umělého. Za prvé je můžeme charakterizovat pomocí adverbialního smyslu, jako to, kde je podstatné, jak došlo k realizaci věci a za druhé v adjektivním smyslu, kde se vyjadřuje, jaká ta věc je.³¹

²⁷ ROMPORTL, 2012. str. 118.

²⁸ Tamtéž.

²⁹ Tamtéž, str. 118-119.

³⁰ Tamtéž, str. 119.

³¹ Tamtéž, str. 124.

3. Turingův stroj a funkcionalismus Hilaryho Putnama

Alan Turing ve své práci *On Computable Numbers, with an Application to the Entscheidungsproblem* představuje teoretický model stroje, který provádí jednoduché, přeprogramované úkony. Přesný popis Turingova stroje je převzat od Filipa Tvrdeho: „Tento idealizovaný stroj má konečný počet vnitřních stavů a obsahuje potencionálně nekonečnou pásku rozdělenou do polí. Pole jsou označena symboly z konečné abecedy, ale pro fungování stroje je možné používat jen dva (např. 0 a 1). Stroj je vybaven čtecí hlavou, která je schopna přečíst symbol na právě analyzovaném poli, a kromě toho umí symboly mazat a přepisovat. Reakce na čtené symboly je definována strojní neboli stavovou tabulkou; po přečtení symbolu se stroj podle předdefinovaných instrukcí rozhodne, má-li se hlava posunout doleva, doprava nebo zůstat na tom samém poli. Chování stroje je tak zcela determinováno jeho konfigurací, tj. současným stavem a právě čteným symbolem. Popsaný stroj je schopen provádět vše, co dokáže jakýkoli reálně existující počítač.“³²

Turingův stroj našel svou aplikaci v logice a ve filozofii matematiky, avšak překvapivě se dočkal popularity i ve filosofii mysli. Hilary Putnam, americký logik, matematik a filosof, byl značně inspirován právě Turingovým strojem, když formuloval svou známou teorii funkcionalismu. Tato teorie vznikla jako protiklad teorie identity představované J. J. C. Smartem. Putnam byl toho názoru, že mentální stavy nelze redukovat na neurální stavy, protože můžeme teoreticky předpokládat existenci bytostí, jejichž mentální stavy jsou realizovány zcela jiným způsobem než u lidí. Podle Putnama nemůžeme tvrdit, že pocit bolesti je ekvivalentní se stimulací určitých nervových vláken mozku, protože mozky různých živočichů se významně liší a navíc mohou bolest cítit i ty entity, které vůbec nemají mozek podobný tomu našemu. Je proto nutné vytvořit takovou teorii, která by tento problém překonala a byla druhově nezávislá. Putnam tak přichází s teorií, podle které mentální stavy nejsou fyzicko-chemické stavy mozku, ale funkcionální stavy celého organismu. Tento klasický funkcionalismus je stavěn na analogii mezi myslícími bytostmi a stroji a proto bývá nazýván strojovým funkcionalismem. Mysl je tedy podobná Turingovu stroji a lidé jsou automaty, které jsou vytvořené z masa a krve. Pokud použijeme současnou počítačovou terminologii, tak naše mysl je software a naše tělo je hardware. U Turingova stroje jsou vstupy a výstupy zajištěny čtením z pásky a zápisem na pásku, zatímco myslící organismy jsou vybaveny sensorickými a motorickými orgány.³³

Není divu, že se funkcionalismus stal jedním z nejpopulárnějších způsobů, jak řešit problém mysl-tělo. Johnem Searlem je právě funkcionalismus označován za ortodoxii současné analytické filozofie mysli. Právě proto si je třeba všimnout určitých odlišností mezi původní Turingovou formulací a Putnamovou reinterpetací. Hlavním rozdílem je Turingovo usilování o teoretické a později praktické sestavení počítače, který by fungoval stejně jako náš lidský mozek. Naproti tomu Putnam se o umělou inteligenci moc nezajímá a je toho názoru, že

³² TVRDÝ, 2014. str. 21.

³³ Tamtéž, str. 21-22.

Turingem popsaný počítač není třeba vůbec sestrojovat, protože ho už dávno máme k dispozici, a je jím právě náš lidský mozek. Rozdíl mezi Turingem a Putnamem je tedy takový, že Turing usiluje o vytvoření umělé mysli, kdežto Putnamovi stačí pouze popsat mysl přirozenou, tedy tu naši.³⁴

³⁴ TVRDÝ, 2014. str. 23-24.

4. Imitační hra

V roce 1950 publikoval Alan Turing svou esej *Computing Machinery and Intelligence*, ve které představil test, který by nám pomohl se rozhodnout, zdali je stroj schopný myslet podobně jako člověk.³⁵ Tento test byl dlouhá léta nazýván „Imitační hra“, název „Turingův test“ se začal používat až od 70. let a jeho zřejmě nejstarší výskyt je v názvu článku „*An Analysis of Turing's Test*“, který napsal James H. Moor roku 1976.³⁶

Turingův test je založen na jakési imitační hře, ve které má stroj za úkol předstírat a přesvědčit pozorovatele, že není strojem, ale že je člověkem. Podle Turinga si máme představit dvě komory, v jedné se nachází člověk a ve druhé je stroj. Stroj se však bude vydávat za člověka. Hlavním úkolem pozorovatele je tedy to, že se pomocí písemného dialogu s testovaným subjektem pokusí určit, ve které komoře je člověk a ve které komoře se naopak nachází stroj imitující člověka. Samozřejmostí pro platný test je i to, že člověk, který je v jedné z testovacích komor, žádným způsobem nepodvádí a nesnaží se pozorovatele zmást tím, že se bude podle odpovědí snažit působit jako testovaný stroj.³⁷

Stroj se tedy musí umět maskovat, aby prošel testem a přesvědčil pozorujícího o tom, že komunikuje s člověkem, nikoliv strojem. Jan Romportl uvádí shrnutí Turingovy koncepce: „Umělá inteligence musí maskovat. Má-li být chování systému napodobujícího činnost nebo myšlení člověka vnímáno jako přirozené, musí systém v interakci se svým okolím stvrzovat určitou formu iluze, případně alespoň neupozorňovat na ty své stránky, jež by v obzorném poli člověka odkryly jeho umělost.“³⁸

³⁵ HAVEL, 2001. str. 34.

³⁶ TVRDÝ, 2014. str. 69-70.

³⁷ HAVEL, 2001. str. 34.

³⁸ ROMPORTL, 2012. str. 121.

5. Námitky vůči imitační hře

Turing ve své eseji zároveň prezentuje námitky, které by proti jeho imitační hře mohly být položeny. Mezi tyto námitky patří: teologická námitka, námitka „hlavy v písku“, matematická námitka, argument z vědomí, argument z různých neschopností, námitka lady Lovelace, argument ze spojitosti nervové soustavy, argument z neformálnosti chování, argument z mimosmyslového chování. Těmito námitkami se budeme následně zabývat.

5.1 Teologická námitka

První námitkou, kterou Turing představuje je námitka teologická: „Myšlení je funkce lidské nesmrtelné duše. Bůh dal nesmrtelnou duši každému muži a ženě, ale nedal ji jiným zvířatům či strojům. A proto žádné zvíře ani stroj nemůže myslet.“³⁹

Turing přiznává, že z této námitky není schopen přijmout ani jednu část. Věří, že zvířata i lidé patří do stejné skupiny, protože v přírodě je mnohem větší rozdíl mezi živým a neživým než je mezi člověkem a ostatními zvířaty. Ortodoxní křesťanský pohled nám bude jasnější, když se podle Turinga podíváme na člena jiné náboženské komunity. Příkladem jsou vyznavači Islámu, kde je jim na základě jejich víry předkládáno, že ženy nemají žádnou duši. Na což mají křesťané samozřejmě jiný názor a upřít duši ženě je pro ně nemyslitelné. Turingovi se zdá, že křesťanské popření toho, že zvířata či stroje nemohou mít duši, znamená vážné omezení boha a je v rozporu s jeho všemocností. Proč by se bůh nemohl rozhodnout, že vytvoří cokoli mu přijde na mysl? V jeho všemohoucí kompetenci by mělo být i vytvořit myslící stroj, což je možné i naším prostřednictvím. Teologické argumenty Turinga moc neohromují, ať už mají podpořit jakékoliv tvrzení. Důvodem je, že tyto argumenty byly v minulosti často neuspokojivé. Jako příklad uvádí žalm, který pojednává o tom, že bůh stvořil Zemi tak, aby se nikdy nehnula. Tímto argumentem, ale i jinými prostředky se křesťané bránili proti Koperníkově teorii.⁴⁰

Filip Tvrďý u této námitky upozorňuje, že Turing, i když to není jeho chyba, operuje s chybně definovanou a možná i nesprávně chápanou všemohoucností boha. Ani v celých dějinách filosofie a přirozené teologie se podle Tvrďého, snad nikdy nenašlo řešení tohoto problému, které by mohlo být uspokojivé z hlediska logické i etické stránky.⁴¹

Tvrďý odkazuje na Turingova životopisce Andrewa Hodgese, který se domnívá, že tato námitka slouží pouze k zesměšnění tradičně scholastického způsobu dokazování v otázkách moderní vědy. Následně nám i Tvrďý představuje svůj možný způsob, jak tuto

³⁹ TURING, Alan. *Computing Machinery and Intelligence*. In: COPELAND, B. J. *The essential Turing: seminal writings in computing, logic, philosophy, artificial intelligence, and artificial life, plus the secrets of Enigma*. New York: Oxford University Press, 2004. str. 449.

⁴⁰ Tamtéž, str. 449-450.

⁴¹ TVRDÝ, 2014. str. 38.

problematickou námitku číst. I když, jak sám přiznává, se tento způsob vymyká původnímu vyznění textu, tak ho považuje za filosoficky produktivnější. Podle jeho výkladu bychom nikdy neměli zatahovat do diskuse jakékoliv pseudoargumenty, které vycházejí z náboženství. Máme si odmyslet z této námitky vše teologické, především boha a nesmrtelnou duši. Pokud to tak uděláme, tak je možné ji redukovat na kritiku Descartova stanoviska. Teologickou námitku, lze podle Tvrdého chápat spíše jako Turingovo odmítnutí karteziánství a nikoliv jako travestii křesťanství.⁴²

5.2 Námitka „hlavy v písku“

Říká se, že když je pštros v ohrožení, tak se nebrání, ale strčí hlavu do písku, protože si myslí, že když nevidí útočnicka, ten ho také nevidí. I když je tato domněnka mylná, tak se metaforicky v určitých případech dá použít i na člověka. Právě proto druhá námitka nese název „hlavy v písku“ a Turing ji začíná slovy: „Následky myslících strojů by byly příliš hrozné. Nechte nás doufat a věřit, že toho nejsou schopny.“⁴³ Tento argument má podle Turinga vliv na většinu z nás, když pomyslíme na stroje, které jsou schopny myslet. Moc rádi bychom se totiž domnívali, že člověk je nějakým způsobem nadřazený vůči ostatním stvořením. A zdá se, že je pro nás nejlepší, když jsme schopni prokázat, že musíme být nutně nadřazení, protože pak nám nehrozí, že bychom mohli přijít o svou privilegovanou pozici. S tímto pocitem, jak Turing dodává je spojena i popularita teologické argumentace. Turing si myslí, že celá tato námitka „hlavy v písku“ není dostatečná k tomu, aby vůbec vyžadovala nějaké vyvrácení.⁴⁴

Filip Tvrdý rozšiřuje původní Turingovu myšlenku: „Dokonce by se dalo říct, že čím menší je skupina *my*, tím větší pocit výjimečnosti zažíváme.“ A následně poukazuje na postupné historické odebrání nadřazenosti člověka. Například Darwin, který zpochybnil naši nadřazenost nad ostatními živočišnými druhy, když naznačil hypotetickou existenci společného předka všeho živého na Zemi. Ve dvacátém století došlo k zrovnoprávnění lidí různých pohlaví, ras a etnicit. „Turingovo teoretizování o myslících strojích je tak útokem na poslední palisádu výjimečnosti, která nám zůstala,“⁴⁵ upozorňuje Tvrdý a dále rozvíjí myšlenku: „pokud jeho názory přijmeme za své a přiznáme myšlení i mechanickým strojům, pak bude do našeho kdysi výlučného *my* patřit kdeco.“⁴⁶ Podle Tvrdého se někteří lidé bojí toho, že by se stroje s umělou inteligencí mohly člověku nejen vyrovnat, ale mohly by jej dokonce i překonat. Poté by bylo jen otázkou času, kdy by si stroje uvědomily svoji nadřazenost a převzaly tak vedoucí postavení ve společnosti. Další obavou je podle něj strach lidí z toho, že budou ve své práci nahrazeni strojem s umělou inteligencí, což může

⁴² TVRDÝ, 2014. str. 38.

⁴³ TURING, 2004. str. 450.

⁴⁴ Tamtéž.

⁴⁵ TVRDÝ, 2014. str. 39.

⁴⁶ Tamtéž.

vést k období ludditského hnutí, které probíhalo v 19. století, kdy luddité ničili stroje v továrnách, protože se domnívali, že jim berou jejich práci. Tvrdý prezentuje možný příklad ze současnosti: „Počítač, který úspěšně projde T-testem, by mohl být ihned zaměstnán třeba v telemarketingu nebo v telefonických poradenských centrech. Ze dne na den by tak přišly o práci miliony lidí, kteří by se mohli považovat právě za oběti nastupující umělé inteligence.“⁴⁷ Tvrdý je toho názoru, že stroje ve skutečnosti lidem práci neberou, ale dávají jim možnost se zabývat různými činnostmi, na které předtím neměli vůbec žádný čas. A stejně tak jako po rozšíření strojové výroby následovala expanze terciární sféry služeb, tak věří, že po nástupu umělé inteligence, by mohlo dojít k něčemu podobnému, například k nárůstu kvartérní sféry vědy a výzkumu.⁴⁸

5.3 Matematická námitka

V úvodu této námitky Turing poukazuje na matematickou logiku, která může být použita k prokázání toho, že existují jistá omezení strojů založených na diskrétních stavech.⁴⁹ Konkrétně má na mysli Gödelův teorém, který ukazuje, že „v každém dostatečně silném logickém systému mohou být formulována tvrzení, které nemohou být prokázána, ani vyvrácena uvnitř systému, pokud není systém sám o sobě nekonzistentní.“⁵⁰ Tvrdý podotýká, že i když se jedná o dva teorémy, tak Turing mezi nimi nerozlišuje, pokud by však používal dnešní přijímanou terminologii, tak by zřejmě mluvil o první větě o neúplnosti.⁵¹

Turing z Gödelova teorému vyvozuje, že existují určité věci, které stroj nemůže udělat. Na některé otázky by stroj v imitační hře odpověděl špatně a na některé by nedokázal odpovědět vůbec, i kdyby měl neomezený čas na odpověď. Turing připouští, že může existovat mnoho takových otázek, které nemohou být zodpovězeny prostřednictvím jednoho určitého stroje, ale věří, že mohou být uspokojivě zodpovězeny jiným strojem. Těmito otázkami myslí ty, na které je možné odpovědět pouze „ano“ či „ne“ a pokládá je za vhodnější, než otázku jakou může být například tato: „Co si myslíš o Picassovi?“. Jako příklad otázky, při jejímž zodpovězení stroj vždy selže, může být uvedena právě tato otázka: „Předpokládejme existenci nějakého určitého stroje, odpoví tento stroj vůbec někdy odpověď ‚ano‘ na jakoukoliv jemu položenou otázku?“ Z toho by měl vyplynout matematický výsledek, který dokazuje omezení strojů, ale toto omezení se netýká lidského rozumu.⁵²

⁴⁷ TVRDÝ, 2014. str. 39.

⁴⁸ Tamtéž, str. 41.

⁴⁹ TURING, 2004. str. 450.

⁵⁰ Tamtéž.

⁵¹ TVRDÝ, 2014. str. 42.

⁵² TURING, 2004. str. 450-451.

Stručnou odpovědí na tento argument je podle Turinga to, že i když je naprosto nesporné, že opravdu existují jistá omezení, která jsou dána jakýmkoliv strojem, tak to neznamená, že stejnými omezeními nemůže být svázán i člověk.⁵³

5.4 Argument z vědomí

Tento argument zakládá Turing na výroku Geoffreya Jeffersona⁵⁴: „Dokud nebude stroj schopen napsat sonet nebo složit koncert pomocí svých vlastních myšlenek a pocitů, a nikoliv pouhým zvolením náhodných symbolů, tak nemůžeme souhlasit, že stroj se rovná mozku, to znamená, že nestačí, když ho jenom napíše, ale musí si být vědom toho, že ho napsal. Žádný mechanismus nemůže cítit (nestačí pouze umělé signály, pomocí jednoduchých triků) potěšení z jeho úspěchů nebo smutek, když se mu spálí výbojka, ani potěšen lichočkami, nemůže být nešťastný kvůli svým chybám nebo být okouzlen opačným pohlavím, ani se nemůže zlobit či propadat depresi, když nemůže dostat to, co chce.“⁵⁵

Turingovi se toto tvrzení jeví jako popření platnosti jeho testu. A podle nejextrémnější formy tohoto pohledu existuje pouze jediný způsob, jak si můžeme být jisti, že stroj je schopen myslet. Tímto jediným možným způsobem je být tím oním strojem a uvědomovat si své vlastní myšlení. Podle tohoto názoru to je stejně aplikovatelné stejným způsobem i u lidí a jediná možnost, jak je možné poznat, že určitý člověk myslí je být tím konkrétním člověkem. Turing upozorňuje, že se ve skutečnosti jedná o vážné solipsistické hledisko. Jako mírně upravený Turingův příklad si dovolím použít, že jsou v blízkosti dvě entity, ta první je označena písmenem „A“ a ta druhá je označena písmenem „B“. Entita „A“ je schopna věřit tomu, že ona sama myslí, avšak popírá tuto možnost u entity „B“. Na druhé straně entita „B“ je také schopna věřit tomu, že ona sama myslí, ale stejně tak jako entita „A“ popírá její mysl, tak i entita „B“ popírá mysl entity „A“. Turing upozorňuje, že místo dohadování nad tímto problémem je obvyklé přizpůsobit se zdvořilé konvenci, že všichni myslí.⁵⁶ Avšak jak nám je známo již z první kapitoly této bakalářské práce, tak tuto zdvořilou konvenci můžeme použít jenom s námi podobnými bytostmi, nikoliv se stroji. Podle Filipa Tvrdeho je pouze jediný způsob, jak s těmito nepodobnými entitami operovat, a tím je vymyslet nějaké kritérium, například imitační hru, díky které budeme moci považovat stroje za určitých okolností za myslící.⁵⁷

⁵³ TURING, 2004. str. 451.

⁵⁴ Geoffrey Jefferson byl britským neurologem a významným neurochirurgem.

⁵⁵ JEFFERSON, Geoffrey. The Mind of Mechanical Man. *British Medical Journal*, 1949, **4616**(1), 1105-1110. str. 1110.

⁵⁶ TURING, 2004. str. 452.

⁵⁷ TVRDÝ, 2014. str. 52.

Následně Turing představuje imitační hru⁵⁸, která je často používaná v praxi, jako ústní zkouška, kde se snažíme zjistit, jestli někdo něco opravdu pochopil anebo jestli pouze „papouškuje“ něco, co se někde dozvěděl:

„Tazatel: První verš vašeho sonetu zní ‚Mám srovnávat tě s krásným letním dnem? Nemyslíte, že by bylo stejně vhodné nebo dokonce lepší použít ‚podzimní den‘?‘

Účastník: Báseň by pak neměla správný rytmus.

Tazatel: A co ‚zimní den‘? To by se rytmicky hodilo.

Účastník: Ano, ale nikdo nechce být přirovnán k zimnímu dni.

Tazatel: Řekl byste, že vám pan Pickwick připomíná Štědrý den?

Účastník: Svým způsobem ano.

Tazatel: Přitom Štědrý den je zimní a nemyslím si, že by panu Pickwickovi toto přirovnání vadilo.

Účastník: To nemůžete myslet vážně. Zimním dnem přece většina lidí myslí typický zimní den, a ne sváteční jako je ten Štědrý.“⁵⁹

Turing po této imitační hře přemýšlí, jak by Jefferson zareagoval na to, kdyby byl stroj píšící sonety schopný odpovědět stejně, jako jsme to viděli ve výše uvedené ústní zkoušce. Neví jestli, by tyto odpovědi považoval za pouhé umělé signály, ale věří, že pokud by byli odpovědi uspokojivé jako ve výše uvedeném příkladu, tak si nemyslí, že by se daly popsat pouze jako jednoduché triky. Zároveň si myslí, že většina z těch, kteří podporují argument z vědomí, mohou být přemluveni, aby toho raději zanechali, než aby kvůli tomu byli nuceni přejít do pozice solipsisty. A poté budou pravděpodobně ochotni akceptovat jeho test.⁶⁰

5.5 Argument z různých neschopností

Tento argument má podle Turinga většinou tuto formu: „Přiznávám, že jsi schopný dokázat, že stroje dokážou všechny věci, které jsi zmínil, ale nikdy nebudeš schopný dokázat, aby alespoň jeden z nich mohl X.“⁶¹ Písmenem „X“ se myslí velmi početná řada věcí a Turing mezi nimi zmiňuje: „Být milý, nápaditý, krásný, přátelský, iniciativní, mít smysl pro humor, odlišit dobré od špatného, dělat chyby, zamilovat se, vychutnat si jahody se šlehačkou, být milován, učit se ze zkušenosti, správně používat slova, stát se subjektem svých vlastních myšlenek, mít mnoho rozmanitých druhů chování jako má člověk, dělat něco skutečně nového.“⁶²

⁵⁸ Pozměněný překlad této imitační hry je převzat od Filipa Tvrdeho, důvodem je, že by doslovný překlad v českém jazyce nedával smysl, kvůli špatnému rytmu. Například v originále je místo „podzimní den“ použito „jarní den“, avšak v češtině má slovo „jarní“ stejný počet slabik jako „letní“ nebo „zimní“.

⁵⁹ TVRDÝ, 2014. str. 33-34.

⁶⁰ TURING, 2004. str. 452.

⁶¹ Tamtéž, str. 453.

⁶² Tamtéž.

Turing věří, že tato tvrzení vznikají pomocí principu indukce. Člověk podle něj ve svém životě už viděl tisíce strojů a z toho jak je viděl, si odvodil několik poznatků, např. že jsou ošklivé; každý je určen pro velmi omezený účel; pokud je po nich vyžadován jiný účel, tak jsou k ničemu atd.⁶³ Avšak v dnešní době se toho dost změnilo a málokdo by řekl například o svém domácím počítači či notebooku, že je ošklivý. Výrobci mnohých značek usilují o to, aby se jejich produkty lépe prodávaly, a v tom nepřekvapivě pomáhá z velké míry i stránka estetická. Samozřejmostí je i nárůst dnešních multifunkčních technologií, u kterých se jejich účel značně rozšířil oproti omezením, která je tížila v Turingově době a je více než pravděpodobné, že k tomu bude docházet i v budoucnu. Problémová část nastane, pokud budeme chtít po stroji, aby posloužil nějakému jinému účelu, než ke kterému byl stvořen. Může se stát, i když je to malá pravděpodobnost, že stroj poslouží i takovému účelu, pro který nebyl původně stvořen, ale je pravděpodobnější, že selže. Tím však není vyloučeno, že by to v budoucnu nemohl jiný stroj zvládnout. I sám Turing osobně si uvědomoval něco podobného a věřil, že limity strojů jsou závislé na paměťové kapacitě strojů, které se v budoucnu budou zvětšovat.⁶⁴

Úsměvná je výtku, která se týká toho, že si stroj nemůže vychutnat jahody se šlehačkou. Protože i jak uvádí Tvrď, to že si na nich někdo není schopen pochutnat, neznamená, že by mu byla ubírána lidskost či inteligence, příkladem je člověk alergický na laktózu.⁶⁵ Další tvrzení o tom, že stroje nedělají chyby, je v dnešní době také vyvráceno a věřím, že se už každému běžnému uživateli počítače alespoň jednou objevilo hlášení o nějaké chybě. Samotný Turing věří, že stroj zvýší svou úspěšnost v imitační hře, pokud udělá chybu například během nějaké matematické úlohy.⁶⁶

5.6 Námitka lady Lovelace

V roce 1822 britský matematik Charles Babbage popsal diferenční stroj a později v roce 1834 analytický stroj. Tyto stroje jsou považovány za první počítače v dnešním slova smyslu. Bohužel Babbage nikdy nedokázal ani jeden ze svých návrhů realizovat. Není jasné, jestli mohli být stroje s tehdejší technologií vůbec zkonstruovány. Babbageův plán byl totiž dost ambiciózní. Jednalo se o kalkulátor, který by byl schopen počítat logaritmické a goniometrické funkce, k čemuž mu sloužila paměť o kapacitě tisíce padesátimístných čísel. Vstup byl prováděn za pomoci dřevných štítků a výstup měl být podle Babbage zajištěn nějakým typem ukazatele nebo tiskárny. Funkci stroje by zajišťovala velice složitá soustava 96 ozubených kol, 24 hřídelí a nespočet táhel, čepů a válců. A podle odhadů by měla být váha stroje až dvě tuny.⁶⁷

⁶³ TURING, 2004. str. 453.

⁶⁴ Tamtéž.

⁶⁵ TVRDÝ, 2014. str. 54.

⁶⁶ TURING, 2004. str. 454.

⁶⁷ TVRDÝ, 2014. str. 55-56.

Francouzský matematik Luigi Menabrea roku 1842 popsal Babbageův analytický stroj a tento text přeložila Ada Lovelace do angličtiny. Anglický překlad opatřila poznámkami, které byly rozsáhlejší než původní text. Její dodatky mimo jiné obsahovaly i popis použití analytického stroje při výpočtu Bernoulliho čísel. Díky tomu bývá lady Lovelace považována za první programátorku vůbec.⁶⁸

Z anglických poznámek od lady Lovelace, konkrétně od poslední poznámky G, Turing převzal další námitku proti svému testu. Ada Lovelace v ní tvrdí: „Analytický stroj nemá žádné domnělé právo na to, aby vytvořil něco originálního. Může udělat pouze to, co mu přikážeme.“⁶⁹ Podle této námitky, platí, že stroj není schopen vytvořit nic nového a pouze dělá to, k čemu ho lidé naprogramují, tedy pouze k věcem, které už lidé znají a ví jak je udělat.

Turing tuto námitku značně pozměňuje do velmi jednoduché podoby, podle které nás stroje nedokážou překvapit. Což podle něho není pravda, protože má plno zkušeností s tím, jak ho téměř neustále překvapují.⁷⁰ Tento argument na námitku lady Lovelace se mi zdá nepřesvědčivý, už jenom kvůli tomu, že Turing přeformuloval původní tvrzení, ve kterém je řečeno, že stroje nedokážou vytvořit něco originálního, do podoby, že stroje nedokážou překvapit. Samozřejmě, že když někdo nebo něco, v tomto případě stroj, vytvoří něco, co od něho nikdo nečekal, tak to vzbudí překvapení. Avšak něco originálního může vzniknout i z toho pokud stroj udělá chybu, pak také dochází k překvapení, ale pouze na základě chyby v systému stroje. Námitku lady Lovelace vidím v tom, že stroji samotnému chybí jeho úmyslná iniciativa vytvořit něco originálního.

Přesvědčivější argument na námitku lady Lovelace, než který uvádí Turing, podle mého názoru prezentuje Filip Tvrďý: „Stejně jako jsou počítače determinovány svým programem a vstupními informacemi, jsou i lidé determinováni genetickou výbavou a sociokulturním prostředím, ve kterém vyrůstají a žijí. Pokud bychom měli tento poznatek vyjádřit žargonem biologie: je determinován genotyp i fenotyp. Náš program je zatím oproti počítačům nesmírně komplikovaný, a proto mají ostatní lidé někdy pocit, že vytváříme něco nového nebo, že je překvapujeme; všechny kousky informace tu už ale byly přítomny i před naším originálním kouskem a předpřipravena byla dokonce i pravidla, pomocí nichž jsme tyto kousky poskládali dohromady.“⁷¹

⁶⁸ TVRDÝ, 2014. str. 55-56.

⁶⁹ MENABREA, Luigi Federico; LOVELACE, Ada. Sketch of the Analytical Engine invented by Charles Babbage... with notes by the translator. Translated by Ada Lovelace. In: TAYLOR, Richard. *Scientific Memoirs (3)*. London: Richard and John E. Taylor, 1843. pp. 666–731. str. 722.

⁷⁰ TURING, 2004. str. 455-456.

⁷¹ TVRDÝ, 2014. str. 60.

5.7 Argument ze spojitosti nervové soustavy

Tento argument spočívá v tom, že naše tělo je tvořeno určitým nervovým uspořádáním, kdežto u strojů to neplatí. Turing podává tuto definici: „Nervová soustava rozhodně není strojem s diskretními stavy. Malá chyba v informaci o velikosti nervového impulsu dopadajícího na neuron může vytvořit velký rozdíl ve velikosti odchozího impulsu. Lze tedy namítnout, že chování nervové soustavy nemůže být imitováno pomocí diskretního stavového systému.“⁷²

Turing s tímto tvrzením souhlasí, avšak není podstatné v imitační hře, protože tazatel z toho nemůže čerpat žádnou výhodu. Důvodem je, že „uspořádání těla“ se žádným způsobem neprojeví na verbálních odpovědích dotazovaného stroje. Stroj by na otázku, která by se týkala jeho „uspořádání těla“, stejně nemohl odpovědět pravdivě, protože jeho primárním úkolem v imitační hře je vydávat se za člověka.⁷³

5.8 Argument z neformálnosti chování

Předposlední argument začíná tvrzením: „Není možné vytvořit soubor pravidel, který by smysluplně popisoval, co by měl člověk dělat v každé myslitelné situaci. Někdo by mohl mít pravidlo, že se zastaví, když uvidí na semaforu červenou. A když uvidí zelenou, tak ví, že může jít. Ale, co udělá, když se na semaforu rozsvítí zároveň červená i zelená? Někdo se rozhodne, že je nejbezpečnější, aby zůstal stát. Ale z tohoto rozhodnutí mohou vzniknout nějaké jiné potíže. Pokusit se stanovit pravidla, aby pokryla všechny možnosti a to i ty, které vyplývají ze semaforů, se zdá být nemožným úkolem.“⁷⁴ Turing s tímto tvrzením souhlasí. Naše výhoda oproti strojům spočívá v tom, že nejsme pevně vázáni, žádnými pravidly a můžeme v dané situaci improvizovat.

5.9 Argument z mimosmyslového vnímání

V poslední námitce se Turing zabývá mimosmyslovým vnímáním, a jeho využitím v imitační hře. Do mimosmyslového vnímání se například řadí telepatie, jasnozřivost či psychokineze. Tyto schopnosti jsou čistě teoretické a nikdy nebyl podán přesvědčivý důkaz o jejich existenci. Turing předvádí příklad imitační hry, kde by se opět tazatel snažil rozeznat člověka od stroje, avšak člověk má tu výhodu, že na rozdíl od stroje ovládá telepatii. Zajímavé je, že Turing vůbec nezdůvodňuje, proč by stroj nemohl být telepatie schopen. Ale přiznává, že pokud bude telepatie jednoho dne možná, tak je nutné zpřísnit imitační hru tak, že by měla být vedena ve speciální místnosti, ve které nepůjde ji použít.⁷⁵

⁷² TURING, 2004. str. 456.

⁷³ Tamtéž, str. 456-457.

⁷⁴ Tamtéž, str. 457.

⁷⁵ Tamtéž, str. 458.

6. Loebnerova cena

Už od těch nejrannějších pokusů používají chatboty⁷⁶ vcelku jednoduché metody vedení konverzace. Jejich algoritmy, které analyzují přirozený jazyk, jsou nastaveny také, aby v každé tazatelově větě hledaly určitá klíčová slova. Tato procedura je známá v počítačové terminologii jako „porovnávání vzorců“.⁷⁷ Jako ukázka může posloužit následující příklad. Libovolný člověk může zahájit komunikaci přes počítač například větou: „Ahoj, od rána mě trápí bolest zubů.“ Program vyhodnotí určitou část, která je syntakticky nejpodstatnější a může na ní vytvořit následující odpověď: „Ahoj, moc mě mrzí, že tě od rána trápí bolest zubů.“

Pokud tato jednoduchá metoda selže, tak je možné, že program odvede rozhovor jiným směrem, za pomoci fráze s nízkou vypovídací hodnotou, kterou může být třeba: „Co se stalo?“ nebo „Jak je to možné?“. Avšak úspěch takového programu není postaven ani tak na složitosti vyhodnocovacích algoritmů, jako spíše na bohatosti databáze s možnými verbálními interakcemi.⁷⁸

Výše uvedený přístup byl vynalezen programátorem Josephem Weizenbauem, který vytvořil úplně prvního chatbota, jenž sloužil jako prostředí pro spouštění zvláštních skriptů obsahujících „osobnost“ účastníka rozhovoru.⁷⁹

Weizenbaum to shrnuje svými slovy: „Napsal jsem počítačový program, s nímž lze ‚konverzovat‘ v angličtině. Lidský partner v této konverzaci napíše na psacím stroji připojeném k počítači svou část konverzace a počítač řízený mým programem analyzuje zprávu, která mu byla poslána, sestaví odpověď v angličtině a zařídí, aby tato odpověď byla napsána na onom stroji.“⁸⁰ Tomuto programu dal Weizenbaum jméno ELIZA. Tímto jménem se inspiroval jednou postavou z *Pygmalionu*⁸¹, konkrétně Elizou Doolittlovou. Věřil, že stejně jako „Pygmalionská“ Eliza, se může i jeho naprogramovaná ELIZA učit „mluvit“ lépe.⁸²

Program ELIZA obsahuje analyzátor jazyka a určitý scénář. Tímto scénářem myslí Weizenbaum sadu pravidel, kterou používají někteří herci, když improvizují. ELIZA může podle určitých scénářů hrát určité role. Můžeme dát ELIZE scénář, díky kterému bude moci být schopna konverzovat například o vaření nebo o bankovníctví.⁸³

⁷⁶ Chatbot je počítačový program, který byl naprogramován tak, aby mohl, v jistých mezích, vést rozhovor s člověkem.

⁷⁷ TVRDÝ, 2014. str. 169.

⁷⁸ Tamtéž.

⁷⁹ Tamtéž, str. 170.

⁸⁰ WEIZENBAUM, Joseph. *Mýtus počítače: počítačový pohled na svět*. Praha: Moraviapress, 2002. Knihovna Ceny Nadace Dagmar a Václava Havlových VIZE 97. str. 8.

⁸¹ *Pygmalion* je divadelní hra, jejíž autorem je George Bernard Shaw.

⁸² WEIZENBAUM, 2002. str. 8.

⁸³ Tamtéž, str. 8-9.

Nejnámějším z těchto „scénářů“ se stal DOCTOR, který za velkého překvapení úspěšně simuloval rogeriánského terapeuta. Volba právě této konverzační role byla naprosto geniální, a to z toho důvodu, že psychologové, kteří se hlásí k odkazu Carla Rogerse, mají v popisu své práce pouze minimálně zasahovat do rozhovoru svými vlastními podněty a slouží jen jako mediátoři sebevyjádření pacienta. Díky tomu, si může program dovolit předstírat, že nemá skoro vůbec žádné znalosti o vnějším světě. Na větu „Včera jsem nechal opravit auto.“ by zareagoval „Povězte mi něco o autě.“, tato reakce by však v jiném kontextu působila nadměrně neobvykle.⁸⁴

Weizenbaum byl šokován, když spatřil, jak u některých lidí dochází k citovému přilnutí k počítači během jejich konverzace s DOCTOREM. Následně uvádí příklad o jeho sekretářce, která věděla, že se Weizenbaum zabývá prací na tomto programu. I když věděla, že se jedná pouze o počítačový program, tak s ním začala běžně konverzovat a po nějaké době dokonce požadovala po Weizenbaumovi aby opustil místnost a dopřál jí soukromí. Weizenbaum uvažoval nad tím, že by systém upravil tak, aby bylo možné kdykoliv zkoumat všechny konverzace, které s DOCTOREM někdo vedl. Poté, co se s touto úvahou podělil s některými lidmi, kteří už s DOCTOREM konverzovali, tak byl okamžitě, jak sám říká, bombardován obviněními, že chce špehovat nejintimnější myšlenky lidí. Weizenbauma překvapuje, že tak jednoduchý program, jako je DOCTOR, dokáže v mnoha lidech vyvolat klamné myšlení, že důvěrně komunikují s člověkem.⁸⁵

Dostí jednodušším, ale o poznání zábavnějším programem se stal PARRY, vytvořený Kennethem Colbym. PARRY imitoval pacienta, který trpí paranoiou. Colbyho volba osobnosti byla obdobně vychytralá, protože většina lidských účastníků, která byla seznámena s psychickým stavem svého protějšku, odpouštěla ochotně některé logické chyby, které byly omlouvány značnou tendencí k přehnané podezřívavosti a patologickými pocity ohrožení. Program tak reagoval na nečekané podněty všelijakými nesouvislými zmínkami o napojení na italskou mafii nebo o podjatém přístupu policie ke slušným lidem. Tazatel tyto reakce většinou interpretoval jako projev pacientovi diagnózy. Programy ELIZA a PARRY jsou od sebe dosti odlišné. V případech, kdy ELIZA používala jakési chlácholivé zdvořilostní fráze, tak právě PARRY byl naopak bez jakéhokoliv důvodu vysoce agresivní a zpochybňoval důvěryhodnost svého oponenta. Mezi jeho reakce například patřilo: „Myslíte si, že jsem lhář?“ nebo „Takoví jste všichni.“ I přes celkovou propracovanost Colbyho simulace umělé paranoie, ji lze těžko považovat za nějaký přínos pro vývoj umělé inteligence. Weizenbaum k tomu ironicky dodal, že bychom mohli navrhnout i počítačovou simulaci autistického člověka, který by neodpovídal vůbec.⁸⁶

První chatboty našly své uplatnění při programování primitivních počítačových her, které se výrazně lišili od těch současných. Tyto primitivní hry, měly pouze textovou podobu. Jejich

⁸⁴ TVRDÝ, 2014. str. 170.

⁸⁵ WEIZENBAUM, 2002. str. 13.

⁸⁶ TVRDÝ, 2014. str. 171-172.

příkladem může být herní série *Zork*, která spadá do žánru „adventure“ a vytvořila ji společnost Infocom. Tato hra měla v 70. a 80. letech minulého století značnou popularitu.⁸⁷

V současné době se u chatbotů využívají totožné postupy, jako u jejich předchůdců. Tyto programy však používají mnohem propracovanější postupy, které navyšují pravděpodobnost k obelstění lidského účastníka, ale dost málo přispívají k pochopení a imitaci lidské inteligence. Bohužel se tento neblahý efekt projevil i ve všech ročnících soutěže o Loebnerovu cenu, která nese své pojmenování po jejím zakladateli a hlavním organizátorovi Hughu Loebnerovi. Soutěž Loebnerova cena byla poprvé vyhlášena roku 1991. Tato událost byla zpočátku velmi prestižní a v organizačním výboru předsedal Daniel Dennett, členem byl dokonce i W. V. O. Quine. Avšak postupem času tato událost začala ztrácet na důvěryhodnosti a byla podrobena velmi intenzivní kritice ze strany filosofů i různých odborníků na umělou inteligenci.⁸⁸

Na základě původních pravidel bylo testováno deset účastníků, z toho byli čtyři lidé a dalších šest byly počítačové programy. Výkon všech účastníků posuzovalo deset soudců. Soudci hodnotili výkon jednotlivých účastníků na škále od 1 do 5. Pokud bylo číslo nízké, tak soudce usuzoval, že daný účastník je stroj, tedy pokud dostal 1, tak to znamenalo, že daný účastník je určitě stroj. Naopak pokud bylo číslo vysoké, tak soudce usuzoval, že daný účastník je člověk, tedy pokud dostal maximální počet bodů, což bylo 5, tak byl soudce přesvědčený, že daný účastník je určitě člověk. Na objektivitu tohoto testu dohlížel nezávislý rozhodčí.⁸⁹

V roce 2003 byla pravidla zásadně revidována a i poté podrobena dílčím změnám. Tím největším zásahem do originálních pravidel bylo rozšíření soutěže o audiovizuální složku, kterou musí daný počítačový program absolvovat po prvním úspěšném složení Turingova testu. V této části musí program zvládnout vést konverzaci, která probíhá v mluvené řeči, to znamená, že v jeho programu musí být zahrnut i prostředek pro analýzu slyšeného jazyka a musí správně reagovat na vizuální podněty, jako jsou obrazy nebo gesta. V této soutěži mohou být uděleny tři medaile: bronzová pro každoročního vítěze, stříbrná pro počítač, který bude soudci považován za člověka v běžné textové konverzaci a zlatá pro počítač, který bude úspěšný v audiovizuální verzi testu.⁹⁰

Pro potřeby Loebnerovy ceny, byl Turingův test omezen, a to mnohými způsoby. Bylo nutné zavést časový limit, který se během vývoje soutěže měnil. Původní délka rozhovoru mezi účastníkem a soudcem trvala kolem patnácti minut, ale postupně byla zkracována na pět minut čistého času. V prvních ročnících soutěže byla přítomna restrikce témat, jejichž soubor byl předem určen. Později však došlo k rozvolnění testu a umožnění komunikovat o čemkoliv. Styl hovoru musí mít podobu přirozené konverzace a soudci mají zakázáno používat jakékoliv léčky a úskoky. V současné podobě pravidel je i obsažen požadavek, aby

⁸⁷ TVRDÝ, 2014. str. 172.

⁸⁸ Tamtéž, str. 172-173.

⁸⁹ Tamtéž, str. 173.

⁹⁰ Tamtéž, str. 174.

byl používán pouze takový slovník, který by byl srozumitelný i pro dvanáctileté dítě. Této omezené variantě bylo často vytýkáno, že restrikce jsou svou povahou dosti problematičké a snižují důvěryhodnost Turingova testu, protože jeho síla původně spočívala v tom, že jsme měli možnost konverzovat libovolným způsobem o čemkoli.⁹¹

Proti soutěži o Loebnerovu cenu vzniklo mnoho kritiky. A není se čemu divit, protože už první ročník ukázal, že počítače velmi špatně komunikovaly a odborníci je dokázali velmi rychle a bez velkých problémů odhalit, na rozdíl od najatých soudců. Mezi nejčastějšími chybami jsme mohli nalézt repetitivní odpovědi na opakující se dotazy, špatně transformovaná zájmena nebo dokonce i naprostá neschopnost reagovat na nesmyslnou sekvenci znaků. Sám Weizenbaum se nechal slyšet, že téměř za pětadvacet let od jeho programu ELIZA nedošlo k nějakému lehce viditelnému vývoji. Dokonce i jeho program PC Therapist, který byl pouze mírnou inovací původní verze DOCTOR, slavil úspěch v prvním, druhém, třetím a pátém ročníku soutěže⁹². Loebnerova cena místo toho, aby informovala o umělé inteligenci, spíše informuje o jednání lidí při jejich interakci s počítači, což ale vůbec není cílem soutěže. Je až nadmíru překvapivé, že i ty nejprimitivnější počítačové programy zvládly oklamat vysoký počet soudců. Podle Tvrdeho z toho zřejmě vyplývá, že se lidé nechají dost snadno zmást, kvůli své snaze k nalezení určité struktury v naprosto chaotických jevech, k naší přirozenosti totiž patří to, že hledáme intenci i tam, kde není, z čehož vyplývá antropomorfizace celé řady nemyslících či neživých entit. Další značně kritizovanou věcí se stala audiovizuální verze testu, která po účastnících vyžaduje schopnosti, které nemají s inteligencí nic společného. V tomto testu by nemohl nikdy uspět člověk, který trpí smyslovým postižením, tedy například slepý nebo hluchý člověk.⁹³

Loebnerovu cenu můžeme paradoxně chápat jako empirickou falzifikaci hypotézy, že sestavení myslícího počítače je v našich momentálních možnostech. I přesto není vůbec třeba ztrácet veškerou naději a oddat se kvůli tomu pesimismu. Dejme lidem ještě šanci a nevzdávejme se, když něco nejde přesně podle našich představ. Sestavení inteligentního stroje je dosažitelným cílem, i když je zatím v nedohlednu.⁹⁴

⁹¹ TVRDÝ, 2014. str. 174.

⁹² Čtvrtý ročník soutěže v roce 1994 vyhrál program TIPS od Thomase Whalena.

⁹³ TVRDÝ, 2014. str. 176-177.

⁹⁴ Tamtéž.

7. Silná a slabá umělá inteligence

Americký filosof John Searle se v úvodu svého článku *Minds, Brains, and Programs* zabývá rozlišením silné a slabé umělé inteligence. Ve slabé umělé inteligenci je hlavním přínosem počítače to, že se při studiu mysli stává velmi mocným nástrojem. Protože nám umožňuje formulovat a testovat hypotézy přísnějším a hlavně přesným způsobem. Avšak podle silné umělé inteligence, není počítač pouhým nástrojem pro studium mysli. Příslušně naprogramovaný počítač se stává opravdovou myslí, v tom smyslu, že disponuje kognitivními stavy. To znamená, že u počítače naprogramovaného silnou umělou inteligencí nejsou programy pouze nástrojem k vysvětlení problému, ale jsou spíše samy o sobě jeho vysvětlením.⁹⁵

Searle zde nemá námitky proti tvrzením o slabé umělé inteligenci, ale zaměřuje se hlavně na silnou umělou inteligenci, konkrétně na tvrzení, že příslušně naprogramovaný počítač má doslova kognitivní stavy a že tyto programy vysvětlují lidskou kognici. Bude se tak zabývat prací Rogera Schanka, protože, jak sám tvrdí, je s touto prací více obeznámen než ostatními podobnými tvrzeními, a hlavně z toho důvodu, že právě Schankova práce poskytuje velmi jasný příklad toho, co chce přezkoumat. Searle zároveň upozorňuje, že všechny následující argumenty nebudou závislé pouze na Schankových programech, ale že by se daly použít například u Weizenbaumova programu ELIZA či u jakéhokoliv Turingova stroje, který simuluje lidské mentální jevy.⁹⁶

Následně Searle stručně popisuje Schankův program. Cílem programu je, aby simuloval lidskou schopnost porozumět příběhům. Pro lidi je charakteristické, že pokud jim vyprávíme nějaký příběh, tak jsou schopni odpovědět i na otázky, které se týkají daného příběhu, aniž by tam byla klíčová informace explicitně vyjádřena. Jako příklad uvádí Searle příběh o muži, který šel do restaurace a objednal si hamburger. Když mu konečně přinesli jeho objednaný hamburger, tak si muž všiml, že je spálený skoro až na uhel. Rozčilený muž odešel z restaurace, bez toho aby hamburger zaplatil nebo nechal spropitné. Teď, když se nás někdo zeptá na otázku, zdali muž snědl hamburger, tak pravděpodobně odpovíme, že ho muž nesnědl. Poté Searle prezentuje druhý příklad, kde stejně jako v prvním příkladě jde nějaký muž do restaurace a objedná si hamburger. Tentokrát, když muž dostane hamburger, tak je s jeho stavem spokojen. Servírce dá velké spropitné, zaplatí a odejde z restaurace. Vám bude položena stejná otázka jako u prvního příkladu, tedy jestli muž snědl hamburger a vy pravděpodobně odpovíte, že ano, muž snědl svůj hamburger.⁹⁷

Schankovy stroje mohou na tyto otázky o situacích v restauraci odpovědět podobným způsobem. Přívrženci silné umělé inteligence, tvrdí, že stroj v této sekvenci otázek a odpovědí nejen simuluje lidskou schopnost, ale je do slova schopen pochopit vyprávěný

⁹⁵ SEARLE, John R. Minds, brains, and programs. *The Behavioral and Brain Sciences*, 1980, 3(3), pp. 417-457. str. 417.

⁹⁶ Tamtéž.

⁹⁷ Tamtéž.

příběh a poskytnout odpovědi na otázky. A že dokonce i to, co stroj a jeho program dělá, vysvětluje lidskou schopnost porozumět příběhu a odpovídat na příslušné otázky. Tato tvrzení přívrženců silné umělé inteligence považuje Searle za zcela nepodložená Schankovou prací. A přichází s tím, že existuje jeden způsob, jak otestovat přítomnost mysli. Tím je zeptat se sám sebe, jaké by to bylo, kdyby naše mysl skutečně pracovala na principu jakékoliv teorie, která může tvrdit: „všechny mysli pracují ,tímto‘ a ,tímto‘ způsobem“. Searle se tedy k zodpovězení tohoto problému pouští do myšlenkového experimentu s „čínským pokojem“.⁹⁸

⁹⁸ SEARLE, 1980. str. 417.

8. Searlův argument čínského pokoje

Na základě minulé kapitoly jsme si mohli všimnout, že Searle nesouhlasí s určitými tvrzeními přívrženců silné umělé inteligence. Základním tématem této kapitoly bude následující otázka: Je schopnost smysluplně odpovídat na jakékoliv otázky známkem toho, že daný stroj se vyznačuje vlastností myslet? K zodpovězení této otázky, jak již bylo řečeno⁹⁹, Searle používá myšlenkový experiment „čínský pokoj“.

Searle uvádí příklad¹⁰⁰, ať si představíme, že stejně jako on nerozumíme ani jedno slovo čínsky a někdo nás zavře do místnosti, ve které se nachází několik košů naplněných čínskými znaky. Avšak v té místnosti se nachází manuál, který je napsán v našem mateřském jazyce a pomůže nám manipulovat s čínskými znaky. V pravidlech je pouze určeno, jak manipulovat se znaky čistě se zaměřením na jejich syntax, ale nikoliv na jejich sémantiku. Jako příklad takového pravidla Searle uvádí: „Vezmi znak škrky-škrk z koše č. 1 a polož ho vedle znaku čmrky-čmrk z koše č. 2.“¹⁰¹

Následně Searle reformuluje tento myšlenkový experiment s čínským pokojem do formy, ve které nám nějakí lidé přidají do místnosti další čínské znaky a zároveň s nimi i nová pravidla pro vydávání znaků z místnosti ven. Znaky, které nám jsou podávány dovnitř, se bez našeho vědomí chápou jako „otázky“ a symboly, které podáváme ven jako „odpovědi“. Searle předpokládá, že náš výkon při manipulaci s těmito symboly bude po nějaké době natolik dokonalý, že bude naprosto nerozeznatelný od odpovědí rodilého mluvčího čínštiny. Takže to nakonec vypadá tak, že sedíme v místnosti, skládáme znaky a vydáváme je ven jako odpovědi na znaky, které nám byly podány.¹⁰²

Je to tedy podobné s Schankovým programem nebo Turingovým testem, kde byl stroj, který dostával otázky a on se podle určitých znaků a programu pokoušel odpovídat jako člověk. Stejně je tomu i u čínského pokoje, kde jsme my na stejné pozici jako stroj u Turingova testu či program v Schankově práci. Také dostáváme otázky a na základě určitého manuálu produkuje odpovědi jako rodilý mluvčí čínštiny.

Searle vidí pointu tohoto myšlenkového experimentu v tom, že z hlediska vnějších pozorovatelů se může zdát, že čínsky rozumíme. Ale ve skutečnosti nerozumíme čínsky ani pouhopouhé slovo. Formálním manipulováním znaků bychom se nemohli nikdy naučit čínsky. A tedy pokud jsme neporozuměli čínsky my, tak nemůže porozumět čínsky ani počítač, protože počítač má k dispozici, stejně jako my, pouze formální program k manipulování s čínskými znaky. Počítač má syntax, avšak trpí absencí sémantiky. Abychom

⁹⁹ Na konci předešlé kapitoly o silné a slabé umělé inteligenci.

¹⁰⁰ SEARLE, John R. *Mysl, mozek a věda*. Přeložil Marek NEKULA. Praha: Mladá fronta, 1994. str. 33.

¹⁰¹ Tamtéž.

¹⁰² Tamtéž, str. 33-34.

porozuměli nějakému určitému jazyku, tak je zapotřebí něco více než jen několik formálních symbolů. U těchto znaků je zapotřebí, aby měly jistý význam.¹⁰³

Teď se pokusím zodpovědět otázku, kterou jsem položil na začátku této kapitoly: zdali je schopnost smysluplně odpovídat na jakékoliv otázky znakem toho, že daný stroj se vyznačuje vlastností myslet. Z toho co jsem již napsal je zřejmé, že se k této otázce staví Searle negativně. A to proto, že strojům neboli počítačům chybí sémantika, mají pouze syntax. Počítače pouze pracují podle určitého programu se znaky, ale těmto znakům kvůli absenci sémantiky nemají žádnou šanci porozumět. A proto nemůže být určité zacházení se znaky podle nějakého programu, bez toho aby daný počítač znal význam těchto znaků, brán za náznak toho, že je schopen myslet.

¹⁰³ SEARLE, 1994. str. 33-34.

9. Eugene Goostman

Cílem mé práce je také zjistit, zdali existuje v dnešní době nějaký software, který by byl schopný zcela projít Turingovým testem ve všech případech s úspěšností 100%. Po přečtení šesté kapitoly o Loebnerově ceně, už může být zřejmé, že bohužel takový program zatím neexistuje. Avšak to neznamená, že by toho v budoucnu nemohl být žádný program schopen. I když podle mého názoru je tato budoucnost značně daleko, tak se umělá inteligence k tomuto cíli blíží alespoň malými krůčky. Důkazem může být program Eugene Goostman, vytvořený ruským programátorem Vladimírem Veselovem.

Tento program totiž exceloval v soutěži, která se konala v roce 2014 na univerzitě v Readingu. Jako první totiž dosáhl celkové úspěšnosti 33% a tím se splnila předpověď Alana Turinga¹⁰⁴, že úspěšnost strojů v pětiminutové imitační hře bude kolem roku 2000 převyšovat 30%. Program Eugene Goostman, se vydával za třináctiletého chlapce, který se narodil v Rusku. Angličtina je tedy jeho druhým jazykem a kvůli svému nízkému věku toho nemůže o světě tolik vědět.¹⁰⁵

¹⁰⁴ Předpověď se splnila, ale s menším zpožděním, u kterého si myslím, že můžeme přimhouřit oko.

¹⁰⁵ UNIVERSITY OF READING. *Turing test success marks milestone in computing history* [online]. 08 June 2014 [cit. 14.3.2017]. Dostupné z: <http://www.reading.ac.uk/news-and-events/releases/PR583836.aspx>.

Závěr

V této bakalářské práci byly představeny pojmy týkající se umělé inteligence a následně byly vysvětleny. Hlavní část byla věnována Turingově testu, na kterém jsme mohli vidět, že pokud stroj umí dostatečně imitovat určité schopnosti člověka, tak může zmást tazatele a ten ho bude považovat za úplně normálního člověka. Následně jsme si zodpověděli všechny námitky, o kterých se Turing domníval, že by mohly být použity proti jeho práci. Většinu těchto námitek se nám úspěšně podařilo vyvrátit. Největší problém jsme měli u argumentu z neformálnosti chování, kde jsme se přesvědčili, že člověk má oproti strojům výhodu, která spočívá v tom, že nejsme pevně vázáni, žádnými pravidly a můžeme v dané situaci volně improvizovat.

Pomocí Searlovo argumentu čínského pokoje jsme zjistili, že i když je stroj schopen smysluplně odpovídat na otázky, tak to v žádném případě není znakem toho, že se daný stroj vyznačuje vlastností myslet stejně jako my lidé. Důvodem je, že na rozdíl od nás, stroje postrádají sémantiku a mají pouze syntax. Počítače se tak řídí pouze podle určitého programu se znaky, ale těmto znakům kvůli své absenci sémantiky nemají naprosto žádnou šanci porozumět. A proto nemůžeme považovat pouhé zacházení se znaky podle nějakého určitého programu za náznak toho, že je stroj schopen myslet.

I přesto, že se mi nepovedlo nalézt, žádný program, který by prošel Turingovým testem s úspěšností na 100%, tak stále věřím, že ve vzdálené budoucnosti to bude možné. Netroufám si říct, za jak dlouho to bude a jestli to pro nás lidi bude mít blahodárný účinek nebo jestli dojde k horší variantě a bude to začátek nadvlády „umělých mozků“, která nás sesadí z naší nadřazené pozice myslících bytostí. Někdo si může myslet, že vývoj v umělé inteligenci může mít v daleké budoucnosti katastrofální následky a může vést k jisté „vzpouře strojů“. Já si však myslím, že pokud jednoho dne dojde k tomu, že umělá mysl bude ve všech svých schopnostech identická s tou naší, tak dojde k tomu, že stroje budou vykazovat i podobně různorodé chování jako má každý z nás. Tím myslím, že je sice možné, že některé stroje se „vzbouří“, ale budou tu i takové stroje, které se rozhodnou stát po našem boku. Avšak uznávám, že to jsou pouhé spekulativní myšlenky a je možné i to, že stroje zůstanou pouze u logického používání určitých znaků.

S nohama na zemi, však věřím, že výzkum umělé inteligence povede k mnoha přínosům. Stroje nám pomocí programů pomohou v různých pracích, v některých nás mohou úplně nahradit a my se pak můžeme zabývat jinými věcmi. Tento výzkum je přínosný i z toho důvodu, že nám může přinést mnoho dalších význačných výsledků ohledně poznání povahy mentálních procesů v naší mysli. Zabývat se umělou inteligencí, je tak podle mého názoru velmi užitečné a věřím, že i přes její počáteční a dá se říci i současné malé krůčky v pokroku, nám postupně ukáže další své cenné výsledky.

Použitá literatura

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

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.

HUME, David. *Zkoumání o lidském rozumu*. Praha: Svoboda, 1996. ISBN 80-205-0521-0.

JEFFERSON, Geoffrey. The Mind of Mechanical Man. *British Medical Journal*, 1949, **4616**(1), pp. 1105-1110.

KRATOCHVÍL, Zdeněk. *Filosofie živé přírody*. Praha: Herrmann a synové, 1994.

MENABREA, Luigi Federico; LOVELACE, Ada. Sketch of the Analytical Engine invented by Charles Babbage... with notes by the translator. Translated by Ada Lovelace. In: TAYLOR, Richard. *Scientific Memoirs (3)*. London: Richard and John E. Taylor, 1843. pp. 666–731.

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.

TURING, Alan. Computing Machinery and Intelligence. In: COPELAND, B. J. *The essential Turing: seminal writings in computing, logic, philosophy, artificial intelligence, and artificial life, plus the secrets of Enigma*. New York: Oxford University Press, 2004. ISBN 0-19-825079-7.

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

UNIVERSITY OF READING. *Turing test success marks milestone in computing history* [online]. 08 June 2014 [cit. 14.3.2017]. Dostupné z: <http://www.reading.ac.uk/news-and-events/releases/PR583836.aspx>.

WEIZENBAUM, Joseph. *Mýtus počítače: počítačový pohled na svět*. Praha: Moraviapress, 2002. Knihovna Ceny Nadace Dagmar a Václava Havlových VIZE 97. ISBN 80-86181-55-3.

Summary

The main objective of this work is to discuss the fundamental problems of artificial intelligence. The main part was dedicated to Turing test. We could see that if the machine is able to sufficiently imitate certain human abilities, so it can confuse the interviewer and it will be considered as human being.

Thanks to John Searle's Chinese Room argument, we found that even when the machine is able to meaningfully respond to questions, so it does not mean that the machine is characterized by the properties of thinking exactly the same as we humans. Machines have syntax but they have not any semantics. Computers are only following a certain program with symbols. But computers have absolutely no chance to understand to these symbols because they have not any semantics. And therefore we cannot be considered mere handling with symbols according to a certain program as an indication that the machine can think.

We could not find a program that could pass the Turing test to 100%. But that does not mean that it should be impossible in future. Artificial Intelligence perhaps goes forward with slow steps, but I think that a lot of great things came from small beginnings.