# Turingův test
## Filozofické aspekty umělé inteligence

Filip Tvrdý

SCHOLIA

*Zpracování a vydání publikace bylo umožněno díky finanční podpoře udělené roku 2014 Ministerstvem školství, mládeže a tělovýchovy ČR v rámci Institucionálního rozvojového plánu, programu V. Excelence, Filozofické fakultě Univerzity Palackého v Olomouci: „Podpora publikační činnosti akademických pracovníků Filozofické fakulty Univerzity Palackého."*



**Turingův test. Filozofické aspekty umělé inteligence**
**Filip Tvrdý**

Vydalo vydavatelství TOGGA, spol. s r. o., Volutová 2524, Praha 5, v edici **SCHOLIA**


Jazyková korektura: Iva Urbanová
Obálka: Lukáš Příbaň
Na frontispisu je fotografie Alana M. Turinga v roce 1951
Photo © National Portrait Gallery, London (artist: Elliott & Fry)
Typografie a sazba z písma Skolar: Dušan Neumahr
Vyrobila: TOGGA, spol. s r. o., Volutová 2524/12, Praha 5
Vydání první, Praha 2014

ISBN 978-80-7476-043-3

# Obsah

# 1 Úvod: problémy s identifikací myslí[1]

Není nic překvapivého na tvrzení, že jsme obklopeni mnoha myslícími, inteligentními entitami. Většina z nás se dokonce s takovými entitami, u kterých jsme s to identifikovat mysl a které považujeme za vědomé, každodenně stýká. Úmyslně přitom prozatím nerozlišuji mezi pojmy „myslet", „být inteligentní", „disponovat myslí" a „disponovat vědomím". V běžném použití jazyka jsou totiž tyto rozdíly jen málo postřehnutelné a vlastně nepodstatné. Těžko si lze představit nemyslící, ale inteligentní bytost nebo bytost, která sice myslí, ale není si toho vědoma.[2] Kdo tedy patří do té již zmiňované třídy entit, které považujeme za inteligentní?

Jsou to především ostatní lidé. Nemusíme se pouštět do žádných komplikovaných filozofických spekulací, abychom přiznali mysl všem, nebo téměř všem lidem. Problematickou podskupinou budou zřejmě někteří mentálně

1 Mé poděkování si zaslouží několik lidí, kteří mi nezištně poskytli podporu. Je to především Iva Urbanová za revizi celého textu a mnoho cenných připomínek k jeho závěrečné podobě. Dále pak Martina Juříková za konzultaci ohledně interpretace Gödelovy věty o neúplnosti, Zuzana Budínská za pomoc s právnickou terminologií, Pavel Šimáček za rady při psaní pasáží týkajících se šachu, Pavel Šuráň za opravu chyby ve Wasonově výběrovém testu a Matúš Šimkovic za upozornění na knihu Tonyho Chemera. Finální verzi textu přehlédli a svými poznámkami doplnili prof. Jan Štěpán, doc. Marek Petrů a doc. Jiří Raclavský. Všechny případné omyly, kterých jsem se při psaní dopustil, padají samozřejmě na mou hlavu.

2 Tímto tématem se budu detailně zabývat v sedmé kapitole, kde se pokusím vyřešit otázky spojené s provázaností inteligence a vědomí. Do té doby žádám laskavého čtenáře o trpělivost.

postižení lidé, řekněme například katatonici, dále lidé v kómatu a pravděpodobně i děti do určité fáze prenatálního či postnatálního vývoje. Ostatním ale mysl rádi přiznáváme, a dokonce si můžeme vybrat z několika filozofických důvodů, proč tomu tak je.

První řešení je z pozic zdravého rozumu, to znamená navazující na nauku o *common sense*, jak ji nalezneme v osvícenské filozofii Thomase Reida nebo analytické filozofii G. E. Moora. Podle této nauky existují tvrzení, o kterých je zbytečné diskutovat, polemizovat s nimi, snažit se je dokázat či vyvrátit. Moore je označoval jako „truismy" a řadil mezi ně širokou škálu tvrzení jako například: 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 (Moore 1925, s. 194). Každé malé dítě či nevzdělaná venkovanka vědí, že ostatní lidé myslí, a proto můžeme toto tvrzení považovat za truismus.

Jiné řešení je pragmatické. To zdůrazňuje, že pochybování o tom, mají-li jiní lidé mysl, je neužitečné. Poprvé se zřejmě implicitně objevuje v Humových *Zkoumáních o lidském rozumu*, v pasáži o oprávněnosti přehnaného skepticismu. Hume tvrdí, že z přehnaného skepticismu – tj. toho, který pochybuje o všem, třeba i o existenci jiných myslí – nikdy nemůže pocházet žádné trvalejší dobro. I když je rozumově nevyvratitelný, přesto selhává v běžném životě, kde „i ten nejzatvrzelejší skeptik je pak k nerozeznání od ostatních smrtelníků" (Hume 1996, s. 216). Podobně argumentuje i o dvě století později Quine, když neméně pragmaticky tvrdí, že je třeba rezignovat na snahu o přímý, nezprostředkovaný přístup k mentálním stavům jiných lidí. Za všech okolností jsme totiž odkázáni jen na vnější, hlavně jazykové projevy ostatních, ze kterých mů-

žeme pouze na základě analogie usuzovat, jaké jsou jejich myšlenky a pohnutky. Tento postoj Quine označuje jako lingvistický behaviorismus a považuje jej za jediné možné stanovisko (Quine 1994, s. 46). Tím, co a jestli vůbec něco se honí v hlavách našim spoluobčanům, si úplně jistí nebudeme nikdy. Přesto je praktické chovat se k nim tak, jako by mysleli a cítili bolest, protože jinak by se mohlo stát, že začnou oni pochybovat o *naší* schopnosti myslet a cítit bolest.

Poslední řešení je nejjednodušší a spočívá v laickém poukázání na neurální podobnosti mezi jednotlivými lidmi. Jsme si podobní, vytvoření ze stejného materiálu, máme stejně uspořádány 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.

Je vidět, že pokud nechceme být malichernými filozofickými kverulanty, nemáme s identifikací myslí u jiných příslušníků lidského rodu větší problém. Věci se ale začnou komplikovat, jestliže se pokusíme připisovat status myslících bytostí za hranicemi našeho biologického druhu. Okamžitě si totiž všimneme, že všechna tři řešení, která jsme vcelku upokojivě aplikovali na případu jiných lidských bytostí, selhávají. Co se týče zdravého rozumu, ukáže se jeho nevyhnutelně nejistá a relativistická povaha. Tvrzení, které jeden člověk považuje za samozřejmé, triviální či banální, může druhému člověku připadat jako neobhajitelné. Moore zřejmě úmyslně vybral sadu zcela nekontroverzních soudů o obvyklých otázkách, lidé ale mají tendenci zašti-

ťovat se zdravým rozumem i v případech, kdy jej aplikovat není možné. Zamysleme se jen nad tím, jak dopadá argumentace pomocí zdravého rozumu v rasových, politických či náboženských otázkách, například ve větách „Lidé všech ras jsou si rovni.“, „Bohatší musejí cítit solidaritu se sociálně slabšími.“ nebo „Křesťanství je nejdokonalejší zjevené náboženství.“ (Upozorňuji, že předkládaná tvrzení nemusí být nutně totožná s přesvědčeními autora této knihy.) Stejně tak existují ti, kteří považují za zcela samozřejmé, že zvířata myslí a snad i rozumí lidské řeči, zatímco pro jiné je představa myslících zvířat, třeba hrochů, absurdní. Pěkně to vyjadřuje starý vtip, podle kterého si filozofové, kteří mají doma psa, myslí, že psi myslí, zatímco filozofové, kteří doma psa nemají, myslí, že psi nemyslí. S odkazem na zdravý rozum si tedy nevystačíme.

Stejně pochybná je i pragmatická argumentace. Viděli jsme, že je užitečné považovat ostatní nám podobné lidi za myslící. U zvířat ale nevíme, co je pro nás vlastně užitečné. Zdánlivě nejužitečnější by bylo považovat je za nemyslící, čímž bychom uchránili před zánikem zoologické zahrady a vyvarovali bychom se globálního vegetariánství, ale stejně užitečná je snaha vědců o hledání pravdy, která může skončit nalezením důkazu o zvířecí schopnosti myslet. Moc nám také nepomůže biologické srovnání nervové soustavy zvířat, protože někteří živočichové jsou nám podobní, ale někteří vůbec ne. To ale neznamená, že by stejné mentální stavy nemohly být vyvolány jinak konfigurovanými nervovými soustavami. Poslední staletí a desetiletí jsou ve znamení rozšiřování práv slabých a ponížených. Výrazně se zvýšila i ochrana práv zvířat, pořád ale žijeme v dosti schizofrenní situaci, kdy za ušlapání ježka může být člověk nepodmíněně odsouzen k dvěma letům vězení (ust. § 203

zákona č. 140/1961 Sb., trestní zákon, ve znění pozdějších předpisů), ale na druhou stranu naše společnost široce toleruje velkochov ustájeného dobytka, klecový odchov drůbeže nebo diskutabilní způsoby porážení hospodářských zvířat. Descartes tím, že radikálně oddělil svět *res extensa* od *res cogitans*, následně zařadil zvířata výhradně do rozprostraněnosti a upřel jim participaci na mentálních stavech, ospravedlnil náš eticky obtížně obhajitelný postoj ke zbytku živočišné říše. V současné době je módní kritizovat tento aspekt kartezianismu a prohlašovat, že aspoň některá zvířata si zaslouží určitý stupeň rovnoprávnosti s lidmi, nedokážeme se ale dohodnout, jaká zvířata a jaký stupeň. Těžko lze předpokládat, že žížala disponuje schopností myslet; na druhou stranu lze stejně těžko tvrdit, že šimpanz ne. Známý aktivista ve věci práv zvířat Peter Singer začíná svou nesmírně vlivnou knihu *Osvobození zvířat* kapitolou o tom, že jsou si všichni tvorové rovni, a končí ji výzvou, aby se každý stal vegetariánem čili jediným eticky ospravedlnitelným typem konzumenta (Singer 2001). Určitou naději pro nás, kteří si ale nechceme upřít občasný hovězí steak, představuje Dennettova teorie mnoha druhů myslí. Podle ní není „mysl" univerzální vlastnost, kterou je entita buď vybavena, anebo není. Ve skutečnosti existuje myslí značné množství. Klíčový je pro Dennetta pojem „intencionální postoj", kterým označuje naši schopnost interpretovat činnost živých organismů i neživých věcí tak, jako by byla záměrná. Intencionalita neexistuje reálně, je jen lidskou kategorií, jejíž pomocí se snažíme odhadnout chování nejrůznějších entit (Dennett 1997).

Všimněme si, že naše manipulace s pojmy, které entitám přisuzují mentální stavy, je značně nepřehledná a nespravedlivá, mohli bychom říct až šovinistická. Někdy totiž

máme tendenci příliš široce aplikovat pojem „mysl“ i na lidi, kteří na to zřejmě nemají právo (například už oni zmiňovaní lidé v kómatu), někdy naopak tento pojem odebíráme zvířatům, která na něj zřejmě právo mají (například vepřům, podle zvířecích měřítek údajně nejinteligentnějším obyvatelům vesnických dvorků). A situace se zkomplikuje ještě víc, pokud se pokusíme do naší teorie identifikace myslí zařadit entity, u kterých je jejich nárok relativně nedávný: stroje, počítače, roboty.

Naše předchozí pokusy o uchopení myslí jsou zcela nepoužitelné. Z pohledu zdravého rozumu nejsou počítače – především ty vypnuté – o nic víc myslící než pluhy nebo bicykly. Pragmaticky o participaci počítačů na mentálních stavech také nestojíme. Proč rozšiřovat naši už tak dosti početnou třídu vlastníků myslí, navíc v situaci, kdy jsou dnes hranice této třídy více než nejisté? A neurální srovnání není možné, protože není co s čím srovnávat. Vždyť naše nervová soustava je s architekturou počítačů naprosto nesouměřitelná, už jen kvůli materiální odlišnosti: naše je založená na komplikovaných organických sloučeninách na bázi uhlíku, zatímco ta počítačová na relativně jednoduchých anorganických sloučeninách na bázi křemíku. S etikou si také moc nepomůžeme. Na rozdíl od roztomilých zvířátek počítače ani v nejmenším nevykazují známky únavy nebo bolesti, těžko u nich hledat stopy sklíčenosti nebo naopak dobré nálady. A ani interpretace pomocí intencionálních postojů nebude příliš filozoficky produktivní, protože bude vždy zatížena obrovskou metaforičností. I ve chvílích, kdy proklínám svůj mobilní telefon, přece vím, že skutečným viníkem nočního vyzvánění není sám mobil, ale rozverný známý. A i když trpělivě přemlouvám svůj počítač, aby nedělal hlouposti a nabootoval operační systém, vím, že se mnohem víc než

o jeho psychický stav jedná o ten můj. Někteří představitelé silné umělé inteligence se nám přesto snaží tvrdit, že *svým způsobem* „myslí" i termostat. Tvůrce pojmu „silná umělá inteligence" John McCarthy ve svém slavném bonmotu říká: „Můj termostat má tři přesvědčení: Že je tady uvnitř příliš teplo, že je tady uvnitř příliš zima a že je tady uvnitř, jak má být." (Searle 1994a, s. 31; překlad Marek Nekula) Tato devalvace predikátů označujících mentální činnost je ale zavádějící a má jen málo společného s otázkou, mohou-li stroje, nebo přesněji řečeno počítače, myslet.

Tato kniha se zabývá nejznámějším pokusem o řešení problému, zdali můžeme počítače zařadit do elitní skupiny hrdých vlastníků mysli společně s většinou lidí, menšinou zvířat a případně některými jinými typy entit, jako jsou třeba hypoteticky existující, biologicky nám zcela vzdálení mimozemšťané. Je to řešení, které na konci čtyřicátých let vypracoval jeden z největších duchů dvacátého století, anglický matematik Alan Turing. Předkládaný text má chronologickou strukturu: nejprve bude vysvětleno vlastní Turingovo stanovisko (kapitola 3), později popsány důležité milníky v diskusi o legitimitě imitační hry. Především se zaměřím na kritické texty Johna Searla, Neda Blocka, Roberta Frenche a Donalda Michieho (kapitoly 5, 6 a 7). Pokusím se ukázat, že ani po více než šedesáti letech od uveřejnění Turingova paradigmatického eseje se nepodařilo najít způsob, jak jeho závěry vyvrátit. Součástí knihy je i popis současných snah o praktickou realizaci požadavků, které na umělou inteligenci Turing nakladl (kapitola 8). Závěr pak tvoří syntetické shrnutí (kapitola 9) a ilustrační ukázka reálného pokusu o imitační hru mezi člověkem a počítačovým programem (kapitola 10).