

Jiří Fiala

LAUDATIO

u příležitosti udělení Ceny Nadace Dagmar a Václava Havlových VIZE 97

JOSEPHU WEIZENBAUMOVÍ

5. října 2002

Zakladatelé vědeckého světového názoru říkali, že filosofové svět stále jen vykládali a vysvětlovali, že však je třeba jej změnit. Svět se jim podařilo změnit k nepoznání, dokonce tak, že jej stále naléhavěji potřebujeme vyložit a vysvětlit. Změny světa dosáhli ovšem tak, že si jej určitým způsobem vyložili a vysvětlili, ale tento jejich způsob výkladu už na vysvětlení světa nestačí. Vyložili si totiž svět tak, že rozumět něčemu znamená být s to s tím manipulovat. Jenže nechtěné a neočekávané důsledky takových manipulací se zpravidla už manipulacím vymykají, takže jim – podle definice – nerozumíme. Označíme je za proto za iracionální, tedy za něco, co není ve správných poměrech a pak nadáváme na poměry. Potřebujeme jiný způsob vysvětlování světa. Odpovědi máme dost, chybí nám však k nim otázky. Nevíme už ani často, na co vlastně odpovídáme. Potřebujeme otázky jiné a nové. Jenže vědecká disciplína spočívá právě v tom, že si některé otázky zakáže klást. Vrcholem disciplíny pak je, když se takovým otázkám už vůbec nerozumí. Je to, jako kdybychom volali na běžce na atletických závodech: „Pane, kam to utíkáte a proč?“

Tato cena je udělována těm, kdo mají odvahu takové otázky klást, ptát se „kam a proč“, a to i s tím rizikem, že se stanou ve svém vlastním oboru disidenty a kacíři. První laureát této ceny, *Karl Pribram*, to vyjádřil stručně: Začnete-li si takové otázky klást, stanete se v očích svých kolegů blázny. Jenže když si je klást zakážete, zbude vám jen to, že se budete dozvídat stále víc a víc o míň a míň.

Kladení takových otázek je však nebezpečné i z důvodu podstatnějšího. Musíme se totiž – jako při zpytování svědomí nebo psychoanalýze – rozpomenout na věci, na které jsme chtěli zapomenout, které jsme vytěsnili. Dozvíme se bolestivě, že některé věci, které jsme pokládali za samozřejmé, samozřejmé nejsou, a že základy, na nichž stavíme náš obraz světa a o něž opíráme své manipulace s ním, nejsou vůbec neotřesitelné a pevné. Uvědomíme si tak jako při zpovědi či psychoanalýze svou zodpovědnost, ale získáme tím odpuštění a svobodu.

Tyto otázky je třeba si klást, i když víme, že na ně žádnou definitivní odpověď nedostaneme. Právě z pokusů o nalezení odpovědí na tyto nezodpověditelné otázky vznikla totiž naše kultura. Druhý laureát této ceny, *Umberto Eco*, to ukazuje na příběhu hledání dokonalého jazyka. Je to příběh naprosto neúspěšný, přesto však jsme se v jeho průběhu dozvěděli mnohé, na co bychom jinak nepřišli. Z vedlejších produktů tohoto hledání povstávaly jednotlivé vědní obory. I počítače a věda o nich, informatika, jsou jedním z nečekaných důsledků tohoto marného hledání.

A konečně pro třetího laureáta této ceny – biologa a filosofa *Zdeňka Neubauera* – je otázka po smyslu světa otázkou ústřední; kniha jeho úvah, která doprovázela loňské udělení této ceny se jmenuje přímo „Smysl a svět“.

Příběh dnešního laureáta Ceny nadace VIZE 97 – profesora Josepha Weizenbauma – je zvláště výjimečný a pozoruhodný. Za prvé se odehrává v oblasti nejexaktnějších a nejtvrděších věd – v logice, matematice a informatice, v nichž přece podle obecného mínění jsou veškeré pochybnosti vyloučeny. Za druhé jeho pochybnosti a otázky byly vyvolány velkým úspěchem jeho vlastního výsledku, který je považován za jeden z nejvýznamnějších kroků na cestě k umělé inteligenci. Za třetí pak tento výsledek – vzdor tomu, jak byl přijímán – odhaluje v čisté podobě povahu té inteligence, kterou počítače mít mohou a boří tak nemilosrdně jednu z iluzí, jeden mýtus.

Vyslechněte tedy, prosím, mé laudatio, chvalořeč, na tento příběh a na jeho tvůrce.

Joseph Weizenbaum se narodil v Berlíně v roce 1923, ale už v roce 1935 rodina emigrovala do Spojených států a unikla tak rasovému pronásledování a vyhlazení. V Americe začal Joseph Weizenbaum studovat matematiku, během války sloužil v armádě, po válce studia dokončil a podílel se na stavbě prvních počítačů na amerických univerzitách. Pro tyto a další počítače navrhoval programovací jazyky (zvláště významný byl jazyk SLIP pro manipulaci s hierarchicky strukturovanými řetězci symbolů, tzv. seznamy), je autorem prvních bankovních a databázových systémů a počítačových sítí. Mimochodem: v roce 1998 byl pozván Goethe-Institutem do Prahy na konferenci „Softmoderna – Literatura v síti“ právě jako tvůrce počítačových sítí.

V roce 1963 byl povolán na Massachusetts Institute of Technology – tedy přímo do chrámu vznikající „computer science“ (informatiky), kde pak působil jako profesor až do svého emeritování.

Od začátku padesátých let se stávalo stále zřejmějším, že počítače jsou schopny nejen „počítat“, ale i zpracovávat texty, manipulovat se symboly, například, že se lze s nimi „dorozumívat“ pomocí speciálních programovacích jazyků. V roce 1950 položil Allan Turing

otázku, zda by se tato schopnost počítačů nedala zdokonalit tak, aby počítače mohly „myslet“, a dokonce zda by nemohly myslet stejně dokonale jako lidé. Cíle by bylo dosaženo, kdybychom nebyli s to rozlišit, zda komunikujeme s člověkem nebo s počítačem. Komunikace by přirozeně musela probíhat v přirozeném jazyku a dílčí otázka zněla, zda počítače dokáží takto komunikovat a zda jsou schopny rozumět našim lidským (nedokonalým) jazykům, mluvit jimi a dokonce nám něco zajímavého a hezkého říci či vyprávět nám nějaký příběh. Joseph Weizenbaum v polovině šedesátých let zkonstruoval takový systém komunikace s počítačem, který téměř splňoval Turingův test: uživatelé rychle zapomínali, že komunikují s počítačem. Tento systém se dokonce dokázal učit mluvit stále lépe. Weizenbaum jej pokřtil ELIZA – podle Elizy Doolittlové, která se také učila mluvit stále lépe, jenže nebylo zřejmé, zda se také stává bystřejší. Podle toho, jaký „scénář“ se ELIZE dal, mohla konverzovat třeba o vaření nebo o řešení školních matematických úloh. Nejslavnější scénář se jmenoval DOCTOR a napodoboval rozhovor s psychiatrem rogeriánského směru.

„Jaký máte problém?“ – „Všichni muži jsou stejní“ – „V jakém smyslu?“ – „Pořád nás s něčím obtěžují.“ – „Můžete uvést nějaký konkrétní příklad?“ – „Ano, poslal mne sem můj přítel.“ – „Váš přítel Vás sem poslal?“ – „Říká, že jsem skoro pořád sklíčená.“ – „Je mi líto, že jste sklíčená.“ – „Je to pravda. Jsem nešťastná.“ – „Myslíte si, že tato návštěva Vám pomůže, abyste nebyla nešťastná?“ A tak dále.

Tato ukázka je ovšem volným překladem, konverzace probíhala v angličtině. Odpovědi počítače bych vlastně neměl překládat, protože počítač nic *neříká*. Je dobré si to říci už teď: počítač své „odpovědi“ ani netiskne, on jen dělá na papír bobky, které si potom *my lidé* vyložíme jako písmena, slova a věty. A počítač si *pamatuje* něco asi tak, jako si papír pamatuje, co jsem na něj napsal. Problém paměti je jiný: vzpomenout si, kam jsem ten papír vlastně dal. Tuto starost papír rozhodně nemá, neumí ani zapomínat, ani si vzpomenout. Právě tak počítače. Takovou metaforikou jim přisuzujeme schopnosti, které nemají. Jinak to asi neumíme, měli bychom si však být stále vědomi toho, co to děláme.

Nějak ale *rozumět* přece musí, bude znít námitka, jinak by přece nemohl takto reagovat. Vůbec ne. Je to něco jakou kouzelnický trik. A platí pro něj totéž co pro kouzelnické triky: to, jak je udělán, bývá mnohem zajímavější než kouzlo samo. Naznačím jen jeden takový trik tohoto programu. Vezměme si třeba větu „Už dlouho jsem velmi nešťastná.“ Počítač s ní zachází jako s *řetězcem znaků* a hledá, zda neobsahuje někde *řetězec znaků* „jsem“. Když jej najde, pak vše, co je před ním zahodí, a zbytek („velmi nešťastná“) si schová. S řetězcem znaků „jsem“ má spojenou tabulku možných začátků odpovědí, např. „Jak dlouho už jste“, „To je mi líto, že jste“, „Řekněte mi něco víc o tom, proč jste“. Jeden z nich náhodně vybere,

doplní schovaným zbytkem a dostane tak odpověď, např. „Řekněte mi něco víc o tom, proč jste velmi nešťastná.“ Jenže co když tam „jsem“ nenajde? Pak najde nějaký jiný vhodně zvolený řetězec znaků s jinou tabulku začátků odpovědí. A když nenajde ani to, pak má připraveny náhradní otázky, např. „Proč si myslíte, že je to důležité?“ Takových „triků“ je tam mnohem víc a jsou velmi pečlivě vyváženy. Spoléhá se přitom na vstřícnost pacienta/pacientky; začne-li konverzace váznout, řekne si, že se asi špatně vyjádřil/vyjádřila. Je to jednoduché – jenže jak se říká „geniálně jednoduché“. Kdysi dávno jsem si ELIZU naprogramoval sám na jednom ze starých velkých počítačů. Byla to jednoduchá verze, dal jsem si tam málo odpovědí. Musím ještě dodat, že to byla verze v angličtině. Konverzace v této jednoduché podobě byla dosti prostoduchá a tak jsem to Elize také řekl – a hezky *česky*: „Eliza je vůl.“ Odpověď, kterou jsem si tam sám před chvíli naprogramoval, mne „dostala“: „We were discussing you – not me.“ – „Mluvili jsme o vás – ne o mně.“

Eliza je i ve své jednoduché verzi velmi působivá a svůdná. I to by mělo být varováním. Profesor Weizenbaum vypráví o tom, jak si jednou jeho sekretářka, která dobře věděla, že jde o pouhý počítačový program, zkusila s Elizou popovídat. Po dvou replikách se obrátila na pana profesora, zda by nešel chvilku za dveře, že se chce Elizu zeptat na něco velmi intimního.

Eliza se stala po svém zveřejnění v r. 1966 velice oblíbenou a začala se šířit s různými scénáři po celém světě. Snadno se předváděla: lidé mají kouzla a hračky rádi. Stalo se však něco, před čím Weizenbaum varoval. Někteří kouzelníci totiž uvěřili, že skutečně kouzelníci jsou, ba dokonce, že nejde o kouzla, ale o realitu.

Prvními takovými kouzelníky byli psychiatři sami. Někteří z nich začali spatřovat v Elize počátek revoluce v psychiatrii, totiž možnost její automatizace. Zdokonalená Eliza by mohla „obsloužit“ (zvláště ve státních nemocnicích, trpících nedostatkem peněz i psychiatrů) za hodinu stovky pacientů. Navíc bude možné pomocí počítačových modelů lépe pochopit podstatu některých duševních nemocí. Za tím se skrývá velmi rozšířený blud: pochopit něco znamená dokázat to modelovat na počítači.

Máme zde jedno z hlavních Weizenbaumových pozdějších témat: otázka nezní, zda něco lze či nelze na počítačích simulovat, např. zda jsou schopny objektivně za nás rozhodovat. V tom problém není, počítač může rozhodovat např. tak, že si „hodí mincí“. Skutečná otázka zní, zda mu rozhodování vůbec máme svěřovat. V principu není problém vybavit počítač plnou „znalostí zákonů“ (jejich doslovným zněním, literou, ne nějakým „duchem zákonů“) a příslušnými odvozovacími pravidly. Počítač by pak mohl vynášet zcela „objektivní“ soudní rozhodnutí. Vyřešil by se tím nedostatek soudů a na několik milisekund by se zkrátila práce

soudů. V tomto případě je snad jasné, že by žádný soudný člověk soudní moc počítačům nepřenechal. V jiných případech však, zdá se, klidně.

Tak stále přetrvává představa, že by počítače mohly rozumět, myslet, mít vědomí. Jistě je to možné, pokud ovšem dáme všem těmto slovům jiný smysl. Když např. inteligencí budeme rozumět schopnost úspěšně procházet testy inteligence, řešit hlavolamy či hrát hry. Pak už je jen krok k tomu, abychom takto zmrzačenými pojmy nahradili pojmy původní. Problém umělé inteligence i slavný Turingův test by tak mohly mít jednoduché řešení: stačilo by počkat, až lidé stykem s počítačovými kouzly zhloupnou natolik, že budou myslet jako počítače. Weizenbaum to jednou formuloval krásně stručně: „Čím hloupější jsou lidé, tím inteligentněji vypadají počítače.“ Jenže taková inteligence by pak už nebyla schopností rozlišovat mezi pravdou a nepravdou, mezi tím, co je správné a co nesprávné. Toto rozlišení za nás udělá někdo jiný a my ani nebudeme vědět kdo, dokonce to už nebude vědět nikdo.

Eliza je stále velmi poučným příkladem: ukazuje v průhledné podobě to, co se stává stále neprůhlednějším a tudíž nebezpečnějším. Právě tato neprůhlednost, anonymita a principiální nezodpovědnost aparátů za počítačovými projekty je podle Weizenbauma tím největším nebezpečím, před nímž chce varovat.

Počítače mohou leccos napodobovat. Mohou simulovat „já“, „vědomí“ i „svědomí“. „Mít“ je ale nemohou, a to z prostého důvodu: neznají bolest a tudíž ani radost. Mohou snad mít čistou inteligenci, nemají však tělo. Nemohou mít víru, naději a lásku.