

Intelligenční exploze

text **CYRIL HÖSCHL**

NALISTUJEME-LI si stranu 220 pozoruhodné knížky Ivana M. Havla a kolektivu jeho 80 gratulantů *Protázky a odvěty*,¹ můžeme se dočíst: „Myslím [...], že ani zdokonalování inteligence (umělé nebo jakékoliv) nepovede ke katastrofě. Někteří počítačovní futurologové mají za to, že až člověk vyvine počítače, které jej svého svou inteligencí předčí (nejenom v ojedinělostech jako dnes, ale ve všem), prokáže tím svou schopnost vyvinout něco intelligenčně lepšího – a tudíž i rychlejšího –, než je on sám. Pakliže počítače budou intelligenčně lepší než člověk ve všem, budou speciálně lepší také v právě zmíněné schopnosti, totiž vyvíjet něco intelligenčně lepšího – a rychlejšího –, než jsou (tentokrát ovšem už) oni sami. Již nebudou muset napodobovat nás, nýbrž sebe samé, a to lépe a rychleji.“

Tuhle kvazilogickou úvahu můžete opakovat a vršit do libovolna. Počítače budou nejen lepší a rychlejší, ale budou se (čili sebe) stále víc zlepšovat a zrychlovat a toto zlepšování a zrychlování jako takové se bude dále zlepšovat a zrychlovat! Dříve či později se utrhnou ze řetězu a dojde k explozi, kterou zmínění futurologové nazývají intelligenční singularitou. Mladší mezi námi se toho prý mají dožít.

Na něco podstatného zapomínají. Za prvé je to ono nenápadné (ale mnou zde zdůrazněné) přesunutí jisté schopnosti člověka (totiž vyvinout něco lepšího, než je člověk) na počítač tak, že se tato schopnost již neorientuje na člověka, ale přímo na počítač – tedy k samotnému nositeli oné vlastnosti. Ten veleschopný počítač by tedy musel o sobě vědět, aby mohl sám sebe zkoumat, potažmo zlepšovat. Vědět o sobě, kde by se to vzalo?

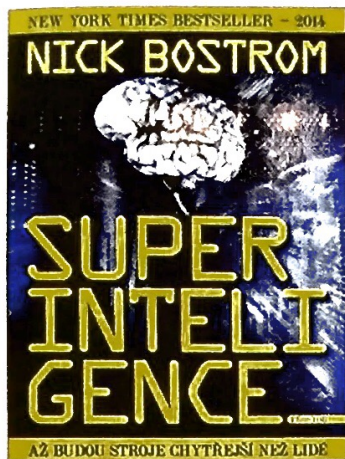
Za druhé, co ani utržením ze řetězu počítače nezískají, je svoboda. Protože nebudou vědět, co to je. Jistě to slovo najdou ve svých mohutných databázích spolu s různými příklady [...]. Ale znát slovo ‚svoboda‘ není totéž jako znát svobodu – slovo přehlédnout lze, ale svobodu nikoliv. (Co je svoboda, přitom dobře – vlastně nejlépe – ví i vězeň, který jí nemá – jinak by netoužil z vězení utéct. Stroj by netoužil.) Bez opravdové, prožívané či prožitelné svobody ani nelze něco pořádně chtít. Proč by ti geniální superinteligentstrojcové měli vůbec chtít sami sebe zlepšovat?“

Tento sice skeptický, ale přesto prozíravý text, který je už jenom douškou za mnohem

starší Ivanovou úvahou o superinteligenci, rezonuje jak s otázkami triviálními, tak s úvahami mnohem komplikovanějšími. Mezi ty triviální patří to, o čem Nick Bostrom mluví jako o metodě izolace: jakýkoliv stroj, který jsme vyrobili, můžeme obrazně řečeno vytáhnout ze zásuvky. To komplikovanější kritické zpochybnění může spočívat v tom, že ve vývoji umělé inteligence se může ukázat, že zákon přechodu kvantitativního v kvalitativní není univerzální a že stroje prostě nikdy správně nepochopí větu „Anička snědla celý talíř“ (budou volat sanitku), větu „ta konvice už vaří“ nebo větu „sedněte si do toho švestkového křesla“. Rovněž tak i při zapojení veškeré fantazie jsou v úvahách o superinteligenci na poměrně tenkém ledě spekulace o emulaci mozku včetně prožívání, jáství, a tedy sémantického uvědomění si sebe sama a dokonce i jednodušších konceptů, jako jsou motivace a pudy. K tomu, aby superinteligentní stroj replikoval sám sebe a sám sebe zdokonaloval nad svou vlastní úroveň, by musel především *chtít* to dělat. V očích skeptiků, jakým je Ivan Havel, je právě ono chtění jedním z vícero obtížně rozlousknutelných oříšků. Nejde přitom jen o oříšky na úrovni individuálních řešení jednotlivých strojů, dokonce ani na úrovni řešení sítí takových strojů, nýbrž i o problém komunikace většího počtu sítí, klonů, klanů a uskupení, jež mohou (?) sledovat všelijaké partikulární zájmy a neustálou kompeticí uvést společenství takové inteligence do evolučního pohybu, jehož zákonitosti jsou zatím jen tušené a důsledky nepředvídatelné. Čtenář by musel být skalním zastáncem panpsychismu a nevidět v životnosti či subjektivitě rozdíl mezi sebou, motýlem a kamenem, aby mohl tyto otázky v případě superinteligence přeskočit. I tak by zůstávala zatím stále ještě na člověku závislá odkázanost umělé inteligence na vnějších zdrojích, a to nejenom energetických, nýbrž i materiálových, takže by počítače při replikaci sebe samých musely organizovat složité těžbu různých vzácných kovů k výrobě

1) Havel I. M. a kol., *Protázky a odvěty*, Lucerna, Praha 2015, 257 stran, ISBN 978-80-7422-362-4.

2) In: A. Pelán (ed.): *Hlavou zeď – úvahy nad civilizací a její budoucností*, dybbuk, Praha 2011, s. 183–217.



**NICK BOSTROM:
Superintelligence
(Až budou stroje
chytřejší než
lidé)**

416 stran, Oxford
University Press,
Oxford 2014, ISBN
980-0-19-873983-8

integrovaných obvodů v odlehlých částech světa, jejich transport a obchod s tím spojený, eliminovat zločin spojený s tímto obchodem, popasovat se s legislativou a organizací dosud nevymřelých společenství atd. atp. Přesto všechno je Nicku Bostromovi nutno přiznat odvahu za to, s jakou vehemencí, komprehenzivitou a odhodláním se pustil do vyčerpávajícího přehledu problematiky, jež futurology, zaměřené na umělou inteligenci a myslící stroje, ovládá přinejmenším od dob Turingova stroje. V očích laika vyvstávají ještě další nezodpovězené otázky. Jednou z nich může být sice nepravděpodobná, ale přece jenom možná konečnost úloh, k jejichž řešení je nutná stále vyšší inteligence. Asi tak jako je konečná plocha na Zemi, je dost možné, že je konečná i úroveň toho, co se na Zemi dá a může řešit a co vyžaduje inteligenci vyšší, než je lidská. Může to znamenat, že naši inteligenci lze sice překročit, ale že ona inteligenci exploze nemusí být de facto explozí a nemusí se vyvíjet ad absurdum.

Dalšími významnými omezeními, jež Bostromova kniha nijak zvlášť nepřipomíná, jsou možné předěly, neřkuli katastrofy v dalším vývoji lidstva, jež mohou dosavadní exponenciální růst zlomit. Může to být válka, opravdu velké stěhování národů, něco, čemu ruský filozof Nikolaj Berďajev říká nový středověk. Ten nepochybně explozi inteligence, jinak též nazývanou singularitou, nezahrnuje. Pokud jde o pokoření člověka, jeho porobení nebo dokonce eliminaci, pak je třeba si také uvědomit, že k ovládnutí mas jsou nezbytné i jiné vlastnosti, než je inteligence či práce s daty. A právě v těchto „charakterových“ vlastnostech ještě dlouho umělá inteligence nebude mít nad člověkem převahu, jestli vůbec. Zasněženější probrání

těchto úvah najde čtenář v publikaci Ivana M. Havla *Cestou k inteligenci singularitě*.² Ostatně onen základní problém, co to vlastně inteligence je a jak se liší inteligence lidská od strojové, stále není uspokojivě vyřešen. Stále ještě jsou stroje designovány k tomu, aby řešily určité úlohy, zatímco člověk je „designován“ k tomu, aby řešil jakékoliv úlohy v jakékoliv situaci. Stále ještě nikdo nevyrobí nejprve univerzálně inteligentní stroj, aby ho pak učil šachy. Podle Havla ona Chalmersova teze, že „stroj, který je inteligentnější než lidé, bude lepší než lidé v konstrukci strojů“, implicitně předpokládá něco, co samozřejmě není: totiž že a) inteligentnější stroj je také lepší v konstrukci strojů, což vůbec nemusí být pravda, a zároveň b) je-li stroj lepší v konstrukci strojů, pak je schopnější konstruovat lepší (inteligentnější) stroje. Podle Ivana Havla (osobní sdělení) zde vystrkuje růžky drobný antropomorfismus, který předpokládá, že nejenom každý inteligentní člověk, jakmile by měl příležitost zvýšit svou inteligenci, by to udělal, ale že by totéž činil i lidsky či nadlidsky inteligentní stroj. Proč by to dělal? A pokud to není antropomorfismus, může to být ještě cosi choulostivějšího, totiž pocit, že obecná inteligence již z definice zahrnuje nejenom schopnost se zlepšovat, ale přímo snahu tak činit. Jako by neexistovala možnost, že někoho něco jednoduše nenapadne - možnost, kterou lze u stroje čekat spíše než snahu něco činit či nečinit. Pokud jde o katastrofické scénáře singularity, pak nás může uklidnit výrok Andrewa Ng, docenta na Stanfordově univerzitě a experta na umělou inteligenci: „Bát se superinteligentních robotů, které by si zotročily nebo by vyhubily lidstvo, je asi jako bát se přelidnění na Marsu. V zásadě je to možné, ale nyní irelevantní.“ I když mezi sci-fi literaturou a vážně míněnou knihou Nicka Bostroma je významný rozdíl, můžeme ji číst se stejným zaujetím jako fikci Freda Hoyla a Johna Elliota *A for Andromeda*. Jinými slovy - čtení je to vzrušující, zajímavé a místy provokativní, ale pro naši současnost ještě nikoli relevantní. Největším nepřítelem člověka totiž stále je - a i po přečtení této knihy ještě dlouho bude - zase jenom člověk. ●