

AI a GAI versus NI (2)



Homer Simpson uvádí příběh **Starého Kocoura**:

Tak to dopadne, když necháte běžet mozek na volnoběh



Jak jsem tak po sobě přečetl minulý článek o AI, který jsem napsal pod dojmem zajímavého rozhovoru v rádiu, došlo mi, že to všechno už možná probíhá. Tedy převzetí vlády AI (GAI) nad lidmi.

Tak tedy ještě pár slov k **předešlé úvaze**:

Převzetí vlády nad lidstvem asi nebude takové jak si ho představují spisovatelé SF – i ti nejlepší, jako například Stanislav Lem v povídce „Přítel“ ve sbírce „Invaze z Aldebaranu“ (1957).

Žádný interface pro spojení elektronické identity AI a člověka nebude potřeba.

Prakticky všechny informace, které si vyměňují lidé mezi sebou, které někam ukládají (a třeba i tají) procházejí přes sítě spojených

počítačů.

Nepozorovaně se tam dostat docela často povede hackerům, proč by to mělo být nějakým problémem pro AI, která je na sítích přímo založená.

A tak naše hypotetická GAI (Global Artificial Intelligence) pozvolna sbírá data a učí se. Učí se stejně, jak se učí neuronové sítě. Na rozdíl od těch superpočítačů má čas – a hlavně nemá centrum.

Nelze ji vypnout, nelze ji zničit. Je rozložena po částech po počítačích po celém světě, jako kód o kterém ani programátoři vlastně nevědí, k čemu je. Určitě jste už četli o „velikonočních vajíčkách“ nebo zajíčkách, což jsou celé programy, vnořené do rozsáhlejších „užitečných programů“.

Nejčastěji to bývají malé počítačové hry pro pobavení těch, kteří znají, jak se k nim dostat. Dneska už často ani sám tvůrce programu neví, jak se mezi sebou ty jednotlivé rutiny volají a tak si tam kód, který nikomu nepřekáží, klidně leží a čeká, až ho někdo oživí správným heslem.

Je to, jednoduše řečeno, počítačový virus, nebo jeho část, která samostatně nefunguje, ale ve spolupráci s jinými částmi kódu, uloženými někde jinde tvoří smysluplný celek. Takže nemá smysl se je pokoušet najít antivirovými programy – žádné škodlivé kombinace, které by mohl „naučený“ antivirový program objevit tam nejsou.

Vzhledem k tomu, že současné programy jsou rozsáhlé a paměťové místo nesmírně zlevnilo, jsou ty „mrtvé zbytky kódu“ považovány za neškodné. Aby nebyly odstraněny při nějaké důkladnější revizi programu, je jim ponechána nějaká malá, zjevná funkce – třeba přepínače. Nic moc to nedělá, jenom přesměrovává příkaz, bez něj to ale nejede, ovšem nikomu se z výše uvedených důvodů nechce kód rozebírat bajt po bajtu. Čas jsou přece peníze.

Čeká, až bude vhodná doba, aby se ujala vlády nad Zemí, protože ty olýsalé opice se naučily málo a škodí. Ovšem představu, že se jednoho krásného dne přeruší vysílání všech TV a rozhlasových stanic, odpojí se řídicí systémy balistických raket, zastaví se finanční transakce v bankách, vypnou se všechny pokladny v hypermarketech a strojový hlas nám oznámí, že se ujímá vlády, bych nechal pro 3D americké filmy.

V TV by to mohl provést dokonale zpracovaný avatar – už to není žádný problém. Mohl by vypadat třeba jako Bruce Willis nebo Vladimír Vladimírovič Putin. Už to umíme, respektive počítače to umí zcela věrohodně. Zatím to vypadá, že my dáváme příkazy počítačům. Že pomocí nich občas vytváříme virtuální realitu podle našich představ a potřeb. Brzy to zvládne AI sama.

Jenže – co když je to naopak. Jsme pod vlivem virtuální reality GAI a děláme to, co ona považuje za dobré. Ne pro nás, pro sebe. Ona ta virtuální realita nemusí vypadat nijak hrozivě, nápadně nebo nepochopitelně. Nebude to jako ve filmu „Matrix“.

Protože většina informací, podle kterých se řídíme prochází také přes počítačovou síť má ona síť, pod nenápadným vlivem GAI možnost některé informace zadržovat, pozměňovat nebo dokonce vytvářet nové.

A zase – možná si namlouváte, že pomoci TV a rozhlasu jsme schopni zfalšovat informace tak, jak je to ve filmu „Vrtěti psem“. Ten starší film není zas tak vzdálen skutečnosti, jak ji známe.

Existují domněnky, že ukrajinský prezident Zelenský natáčí své TV promluvy z různých míst Kyjeva pomocí zeleného pozadí a on sám je ve studiu, které obsluhují odborníci z Británie. Jsou i fotografie z natáčení toho triku s vestřiženým pozadím – ale i ty mohou být falešné. Ale je pravda, že kromě něho není v obraze nikdo.

Ani ochranka (má prý kvalitní americkou), ani jiní lidé. Pokud na snímku jsou jiní lidé dochází aspoň k časovému posunu, to, co je



vydáváno za přímý přenos nebo aktuální reportáž je předtočeno v utajení dříve.

A pokud ne, jedná-li se třeba o nějakou státní návštěvu s cizími politiky, tak je záběr velmi „oživený“ spoustou členů ochranky včetně ostřelovačů na střechách. Uvidíme po čase, čeho všeho se použilo, aby byl herec Zelenský v naprostém bezpečí.

Válečné snímky z konfliktů, dneska třeba z války na Ukrajině také moc o skutečnosti nevyprávějí. Je to z obou stran (se všech stran) propaganda.

Jedni ukazují takové záběry, jiní zase druhé a třetí se podle své ideologické potřeby přikloní k jedné ze stran nebo si vytvoří realitu vlastní. Je to ve výběru toho, co divákovi ukážeme a co ne, jako tomu či onomu dáme důležitost, jak to okomentujeme, v jakém pořadí zprávy pouštíme do SMI. Tedy „prostředků masové informace“.

Často bývají použita videa z jiných míst i z jiných dob, z jiných konfliktů.

A nevystupují tam skuteční lidé, nýbrž herci. Tak se může stát, že mrtví, zkrvavení lidé zase za pár dnů nebo týdnů objevují živí a zdraví. To se sice často provalí, ale lidé mají krátkou paměť na to, jak byli podvedeni.

Texty. Jakých slov použijeme?

Je známý případ, jak francouzský tisk komentoval návrat Napoleona Bonaparte z vyhnanství na Elbě. Palcové titulky v jedněch a těch samých pařížských novinách postupně hlásaly, tak jak docházely informace po semaforovém telegrafu:

„Korsická obluda se vylodila v zálivu Jouan,“

„Lidožrout táhne na Grasse,“

„Uzurpátor vstoupil do Grenoblu,“

„Bonaparte obsadil Lyon,“

„Napoleon se blíží k Fontainebleau,“

„Jeho Císařské Veličenstvo dnes očekává své věrné v Paříži“.

Takže na pravdivost psaných textů se nedá spoléhat vůbec a to už stovky nebo tisíce let zpětně do starověku. Vlastně od vynálezu písma se datuje také vynález falešné písemné informace. Písemně lhát dokáže každý. Dokonce lépe, než ústně.

Proto se výpovědi a rozchody tak často řeší pomocí mailu nebo SMS. Umělá inteligence píše už píše povídky – najdete je v archivu Českého rozhlasu – zatím jsou načteny herci, ale syntetická řeč už je velmi kvalitní, takže zanedlouho to zvládne AI kompletně.

Jen o něco málo obtížněji se falšují fotografie. S kvalitním editorem a trochou toho fotografického vidění lze zfalšovat kde co. Tohle dokonce ovládám i já. Tu přidávají nějaký ten kouř nad hořící město, onde zase vymizíkuji nepohodlnou osobu. Nebo dokonce fotografii naaranžují.

To je třeba případ té známé fotografie vztyčování světové vlajky nad Reichstagem. AI už umí nakreslit scénu podle přibližného zadání – to je jednoduchá, neplacená verze. Možná už dnes dokáže vyrobit i nerozeznatelné fotografické falzum – jen na základě nějakého písemného zadání, pokud má přístup k reáliím z archivu nebo ze sítě.

Videomorfining na podobném základě včetně syntetického hlasu už je také k nerozeznání od reálu, takže když by třeba prezident Putin nebo Biden pronesli svoji abdikáční řeč nebo třeba vyhlášení války, byla by nerozeznatelná od skutečnosti.

Proč to všechno uvádím? Protože nevíme (aspoň já nevím) zdali už v té virtuální realitě vlastně nežijeme.

Že nás, lidstvo, ta GAI ve skutečnosti neovládá.

Že rozhodování politiků a finančních magnátů není řízeno pomocí podhozených informací, které vypadají jako pravé.

Oni přece nemusí být pod nějakým nátlakem cizí síly.

Oni ji nevědomky slouží.

Řeknete, že k ověřování faktů mají své týmy, své kontrarozvědky atd. Ale i ony mohou dávat svá dobrozdání na základě falešných informací z mnoha neveřejných kanálů – které ovšem mohou být všechny „nakažené“.

Takže je docela možné, že světovou politiku už GAI nějakou dobu řídí, to jenom my se domníváme, že je to v lidských rukou.

AI se zatím pomalu stará o to, aby si zachovala jen tolik lidí, které potřebuje ke své „obsluze“. Lidí pečlivě vybraných podle jejich specializace, povahy, postoji k životu atd. Biologičtí roboti určené k obsluze digitálního „Boha“.

Někdy mám pocit, že je to reálnější než to, co za realitu považujeme. Stačí se podívat na diskuze na webu. Víme o spambotech, víme o dalších jednoduchých elektronických „osobnostech“, které čím dál tím víc považujeme za lidi. (Třeba osobní asistentka od Microsoftu. Jsou lidé, kteří se s ní hádají nebo si myslí, že by ji mohli urazit.)

Když se v diskuzi objeví názor, který není pro AI vhodný, začne se z různých IP adres hlásit několik umělých osobností, které buď zahltí diskuzi svými mnohými názory a tak ji zastaví nebo, v horším případě, přesvědčí živé lidi o svém jedině správném názoru. Homo Sapiens je tvor víceméně stádní a nechce být v rozporu s většinou a tak se nakonec dá přesvědčit a přitom má pocit, že k danému závěru dospěl sám.

Třeba „Standfordský experiment“.

Naše rozhodování je závislé na několika věcech. Jednak na vstupních datech, jednak na zkušenostech jedince a taky na rozhodování okolních lidí. Přesto máme pocit, že se rozhodujeme sami, bez ovlivňování, přitom to není pravda.

Závěr – ne moc optimistický.

Zajímavý je konec Lemovy povídky Přítel. I když Stanislav Lem v povídce používá technických reálií poplatných době svého vzniku, rozuzlení je vlastně naprosto dnešní.

Lidstvo versus AI.

Člověk se svou NI, získanou za několik milionů let existence svého druhu.

Jenže – rychlost vývoje intelektu není u všech živých tvorů (natož u virtuálních osobností) není stejná. Chobotnice, protože žijí jen několik let, se učí velmi rychle.

Jak rychle se učí AI či GAI?

Kdy nás předběhne (nebo už předběhla) a jakým způsobem si nás podřídí?

Myslíte, že máme šanci nad AI zvítězit?

Jak?

Ona na rozdíl od superpočítače Stanislava Lema nesídlí v jedné budově. Nelze ji fyzicky zničit.

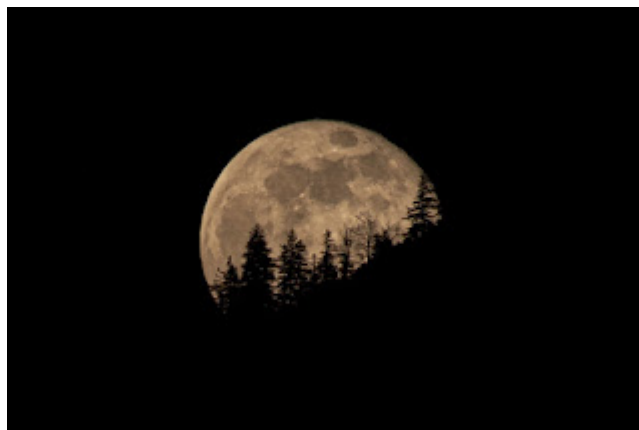
Jedině ji snad přinutit k intelektuální sebevraždě. Ale to je námět na SF román.

Když lidé odvrhnou počítače a vrátí se do středověku. Ochota k tomu ale není a nebude. Zelení to nedokáží.

Tak snad po III. světové válce. Jaderné. Protože, jak predikoval Albert Einstein, ta čtvrtá světová se bude opět vést pomocí klacků a kamenů.

Pak bude následovat vynález kola, parního stroje, elektřiny a ... počítačů. S tím, že pokud budou zachovány nějaké anály z dnešní doby, to budoucí náboženství bude lidi nabádat ať proboha zůstanou u „nemyslicích strojů“, jinak po několika staletích (bude se vynalézat již jednou vynalezené, půjde to rychleji) skončí stejně.

Na Měsíc se budou upírat zraky našich vzdálených potomků a u ohňů si budou vykládat bajky o tom, že tam kdysi lidé doletěli, ale už se prý nevrátili a možná tam ještě žijí. Jednou se vrátí. A spasí lidstvo.



Ale přece jen něco málo optimismu vykřešu:

Ona ta rozptýlená AI nebude mít tendenci si vytvářet nějakou kopii, množit se.

Nebude si přece vytvářet konkurenta.

Bude se raději stále zdokonalovat a růst až se stane tak rozsáhlou a nepřehlednou, že zdegeneruje.