Tomáš Mikolov: Každá technologie se dá zneužít

Tomáš Mikolov, laureát Ceny Neuron, patří k předním českým expertům na umělou inteligenci a neuronové sítě. Tomáš pracoval pro technologické giganty jakými jsou Google nebo Facebook a je jedním z vědců, kteří stojí za zdokonalením Google překladače.

Ještě jako student vyvinul Tomáš modely jazyka založené na rekurentních neuronových sítích a způsobil tím obrat ve svém oboru. Rekurentní neuronové sítě dokážou totiž popsat strukturu jazyka mnohem přesněji než jiné předchozí přístupy. Ptali jste se Tomáše Mikolova na Facebooku na vše, co vás zajímá z oblasti umělé inteligence. Projděte si jeho odpovědi.

Jak se díváte na problém s kontrolou rozhodování a chování umělé inteligence a strojového učení. Lze nějakým způsobem do budoucna řešit a kontrolovat, že software dělá jen to, co má a v případě chyby zjistit, proč tu chybu udělal? Nebo je to něco neřešitelného, "by design" a zůstane to jedním z rizik použití těchto technologií?

Explainable (vysvětlitelná) AI je jeden ze současných směrů výzkumu, který by měl vést na modely, u kterých bude více jasné, proč dělají jaká rozhodnutí. Nicméně je dobré zmínit, že různá prohlášeni typu "vytvořili jsme úžasnou AI, ale nikdo neví, jak vlastně funguje" jsou pouze marketingová hesla. Vědci ví naprosto přesně, jak jejich modely fungují a když je třeba, dokážou zjistit, proč ke kterému rozhodnutí došlo.

Ta nejistota spíše vychází z toho, že v obrovském statistickém modelu s miliony parametry není možné jednoduše říct, co který parametr sám o sobě znamená - ale ten stejný problém budete mít i třeba u složitých modelů pro předpověď počasí. Pokud tedy řešíte úlohu, kde vám stačí třeba 95 % přesnost, pak je strojové učení vhodné. Pokud potřebujete 100 % přesnost, tak je lepší (pokud je to možné) použít klasické programovaní.

Zajímám se o téma AI a jsem učitelka ICT. Co by se podle Vás měli žáci dozvědět o tomto tématu v rámci základního vzdělávání na ZŠ?

Myslím, že některé z AI přístupů k řešení problémů by se mohly vyučovat v rámci matematiky a programovaní už na ZŠ. Příkladem mohou být jednoduché klasifikátory učené z dat, nebo třeba algoritmus Monte Carlo. Tyto jsou nejen snadné na pochopeni, ale dají se uplatnit docela dobře i v praxi.

O které oblasti lidské činnosti, kde se AI využívá, byste žáky informoval nejdříve? Lze odhadnout, ve kterých profesích se tyto děti, které budou v práci za 5 až 10 let, s AI (ne) potkají?

Určitě se s AI algoritmy setkají ti, kteří budou ve své profesi používat počítače. To můžou být například doktoři (lepší a přesnější diagnóza a návrh léčby s pomocí AI), programátoři (odstraněni spamu, strojový překlad, doporučováni obsahu), učitelé (nástroje pro personalizaci výuky a opravovaní testů) atd.

Mohl byste popsat nějakou konkrétní aktivitu pro ZŠ, kterou zmiňujete výše a máte s ní zkušenost?

S výukou na ZŠ zkušenosti nemám, nicméně věřím, že právě třeba Monte Carlo je hezký algoritmus, který lze snadno vysvětlit už na prvním stupni ZŠ. Případně v hodinách programovaní je možné použít některé z AI nástrojů třeba na klasifikaci obrázků, nebo třeba word2vec algoritmus na shlukovaní slov, která se vyskytují v podobných kontextech (což může být zajímavé v hodinách češtiny na druhém stupni).

Jak pohlížíte s ohledem na rozvoj technologií a digitalizace (zejména státních institucí) na otázku bezpečnosti a svobody. Narážím tím na budoucí možnost lepší státní kontroly občanů, což na jedné straně sice může vést k lepší identifikaci „hříšníka“, ale v případě špatně nastavených norem ve společnosti, pak může dojít až k totalitně fungujícímu státnímu celku.

Každá technologie se dá zneužít a jak se říká, cesta do pekel je dlážděna dobrými úmysly. Nakonec i dnes vidíme, že v některých zemích dochází k masivnímu sledování obyvatelstva za účelem zpomalení šíření viru (před tím to stejné bylo ve jménu boje proti terorismu). Otázka je, k čemu to v budoucnu povede, pokud dovolíme státním institucím maximalizovat svoji moc.

V současné době dokážou metody hlubokého učení neuronových sítí úžasné věci. Ať už se jedná o rozpoznávání, syntézu nebo transformaci jazyka, obrazu i zvuku. K těmto úkonům však bývá potřeba obrovské množství vstupních tréninkových dat. Existují nějaké nadějné směry principiálního zefektivnění tohoto učení? Případně metody abstrakce obecných zákonitostí z množiny tréninkových dat?

O toto se dlouhodobě snaží řada vědců – jedná se o problém učení/generalizace, kdy ze stejného množství trénovacích dat chceme vytvořit modely, které budou lépe fungovat na neviděných (testovacích) datech. Jeden ze směrů je i začlenění modelů fyziky (například v rozpoznávaní obrazu/videa), nicméně toto je stále,

pokud vím, nevyřešené. Jinak tech směrů je více, například kombinace supervizovaného a nesupervizovaného učení, také *reinforcement learning* (kde počítači jen sdělujeme, že udělal něco dobře/špatně, ale už ne co měl přesně dělat) atd.

Do jaké míry myslíte, že je pro vývoj obecné umělé inteligence potřeba její propojení s fyzickým světem? Mám na mysli např. zda by na její natrénování stačil přístup k rozsáhlé databázi, případně celému internetu. Nebo zda je potřeba napojení na nějakou soustavu senzorů a manipulátorů pro fyzickou interakci se světem a zpětnou vazbu.

Obecnou umělou inteligenci dnes nikdo vytvořit nedokáže (a to ani zdaleka), takže mužů uvést jen svůj dohad. Myslím, že by mělo stačit propojení s naším světem na skoro libovolné úrovni (tj. i přístup k internetu bez fyzického těla/robota se senzory by měl stačit). Na druhou stranu si myslím, že bude netriviální vytvořit obecnou umelou inteligenci, které budeme rozumět – tomuto se věnuji vice v článku *„A Roadmap towards Machine Intelligence“* z 2015.

Do jaké míry je v současnosti rozpracováno řešení tzv. kontrolního problému? Tedy otázky, jak zajistit, aby cíle případné obecné a samostatně se učící/zlepšující umělé inteligence zůstaly sladěny se všeobecnými zájmy lidstva a aby nepřesnost či přehlédnutí v jejich počátečním nastavení nevedla k nějakým neočekávaným negativním důsledkům?

V současné době jsme velmi daleko od vytvořeni samostatných AI systémů. Myslím, že projektů, které navrhují různá řešení, jak kontrolovat AI je řádově více než projektů, které opravdu mohou vést ke vzniku obecné AI. V článku, který zmiňuji výše navrhujeme vytvořit AI, která bude velmi úzce spjatá se svým uživatelem, tedy svým způsobem bude naší součástí (podobně jako auto je svým způsobem naši součásti, když v něm sedíme a řídíme ho).

Zajímalo by mě, jak se stavíte k otázce umělého života. Jaké atributy by měl mít například chatbot, abychom ho mohli už považovat za živého?

Za živého bych snad považoval chatbota, který má (víceméně) neomezený potenciál v tom, co se dokáže efektivně naučit (tj. podobnou rychlostí jako člověk, nebo vyšší). Program typu 'If (vstup_uzivatele == "Ahoj") print ("Nazdar, jak se máš");' za živý určitě nepovažuji; a přitom dnešní nejlepší chatboti vypadají právě takto, tedy jako databáze přednastavených odpovědí s nulovou schopností se učit/adaptovat.