

Prvý veľký stret s umelou inteligenciou sme vôbec nezvládli, vraví informatik Halgaš a odporúča päť kníh o AI

[Filip Struhárik](#) 6. marca 2024 9:49



Tento text načítal neurálny hlas. Najlepšie sa počúva v [aplikácii Denník N](#), aj s možnosťou stiahnutia na počúvanie offline. Našli ste chybu vo výslovnosti? [Dajte nám vedieť](#).

Umelá inteligencia už nie je len zápletkou v sci-fi príbehoch, ale súčasťou našich každodenných životov. Ako sa vyvíja, aké má nedostatky a ako zmení svet? O tom sme sa rozprávali s Tomášom Halgašom. Umelú inteligenciu (AI) študoval na univerzite v Oxforde a dnes vybral päť kníh, ktoré sa tejto téme venujú a ktoré sa podľa neho oplatí prečítať.

V rozhovore vysvetľuje, že ľudský mozog má stále väčšiu výkonnosť ako najväčšie počítače. „Prebiehajú v ňom oveľa komplikovanejšie procesy. Každý neurón v mozgu je oveľa zložitejší ako umelý neurón, ktorý vieme naprogramovať. V počítači však vieme urobiť výpočty oveľa rýchlejšie,“ vraví.

Halgaš varuje, že pre umelú inteligenciu môže ľudstvo čeliť aj existenčnému riziku. „Umelá inteligencia je totiž extrémne mocný nástroj,“ povedal. Prvý veľký stret s umelou inteligenciou využívanou na sociálnych sieťach sme podľa neho vôbec nezvládli, no napriek tomu zostáva mierne optimistický.

Päť kníh o umelej inteligencii



- [Artificial Intelligence: A Modern Approach](#) (Peter Norvig, Stuart Russell)
- [How to Create a Mind](#) (Ray Kurzweil)
- [The Alignment Problem](#) (Brian Christian)
- [Život 3.0](#) (Max Tegmark)
- [Spravedlnost: Co je správné dělat](#) (Michael J. Sandel)

Umelú inteligenciu ste študovali na univerzite v Oxforde. Ako k tomu došlo?

Bol som pomerne šikovný študent, taký typický jednotkár. Robil som aj rôzne matematické súťaže a takisto ma veľmi bavilo programovanie. Najprv som hrával počítačové hry, potom som ich

upravoval a nakoniec ma už ani nebavilo ich hrať. Len som sa bavil na tom, čo všetko viem zmeniť a vymyslieť.

Keď som si mal vybrať univerzitu, učitelia aj rodičia mali odo mňa veľké očakávania. Dostal som vtedy takú knižku, volala sa Good University Guide, a ako najlepšia univerzita v nej bol uvedený Oxford. Tak som si povedal, že to skúsím. Samozrejme, prihlášky som si poslal aj na iné univerzity, ale chcel som skúsiť aj tú najlepšiu. Nakoniec to vypálilo tak, že ma prijali.

Vy ste najprv nešli študovať umelú inteligenciu, ale kybernetickú bezpečnosť, však?

Išiel som študovať matematiku a informatiku a prihlásil som sa na „college“, na ktorej bolo viacero expertov na kyberbezpečnosť. Veľmi ma to bavilo a myslel som si, že to budem študovať. No keď som objavil témy ako strojové učenie a ako môže fungovať umelá inteligencia, tak ma to úplne vtiahlo.

Čím vás umelá inteligencia tak oslovila?

Dovtedy všetko, čo som programoval alebo upravoval, som musel riadok po riadku vlastnoručne nakódovať. No zrazu tu bol algoritmus, ktorý som vedel naprogramovať, aby sa učil sám. A potom to vedel robiť, a to často ešte lepšie ako ja.

Ako vyzeralo štúdium umelej inteligencie na Oxforde?

Oxford má veľa modulov, ktoré si človek nevyberá, ale sú povinné: analýza, algebra, zostavenie počítačov, výpočtové modely... V druhom ročníku si študent začína vyberať nejaké predmety, v treťom a štvrtom už študuje to, čo chce. Vtedy si už hľadá nejaké zameranie.

Ja som v druhom ročníku robil nejakú základnú umelú inteligenciu, v treťom a štvrtom ročníku som si vybral všetko, čo som o umelej inteligencii našiel. Ale dal som si aj skvelé predmety, ako kvantové počítače, lebo to bolo nesmierne zábavné.

Oblasť umelej inteligencie sa momentálne veľmi rýchlo mení. Aké to bolo študovať? Máte pocit, že to, čo ste v škole preberali, je stále aktuálne?

Bolo to skoro pred desiatimi rokmi. Mám dojem, že akurát vtedy, keď som bol na univerzite, sme prichádzali do novej éry umelej inteligencie. Spomínam si napríklad, že keď sme sa učili procesovanie prirodzeného jazyka, tak najsilnejší model bol ešte napísaný podľa pravidiel pravopisu. Celý algoritmus bol teda napísaný ručne: teraz hľadaj prídavné meno, teraz hľadaj podstatné meno a snaž sa z toho niečo vyvodiť. My sme boli asi posledný ročník, ktorý sa to takto učil.

No mali sme už aj prvé lekcie od profesorov, ktorí pracovali v DeepMinde (*Spoločnosť, ktorú dnes vlastní Google. Spoluzakladateľ firmy Mustafa Suleyman vlani vydal knihu [The Coming Wave](#) – pozn. red.*). Boli to priekopníci, ktorí nám ukazovali, že sa to dá aj úplne inak.

Keby som teda šiel teraz na Oxford, učil by som sa to iným spôsobom. No získal som pevné základy. Základné princípy, na ktoré je na univerzitách čoraz menej času, som sa ja naučil. Mám teda pocit, že mám vďaka tomu oveľa lepšie chápanie, ako veci fungujú.

Počul som, že keď sa objavil ChatGPT, zavládla medzi študentmi informatiky dezilúzia. Mysleli si, že šli študovať odbor, ktorý je veľmi dobre platený, perspektívny, a že programátori budú na pracovnom trhu nenahraditeľní. A zrazu zistili, že nástroje umelej inteligencie vedľa mnohých programátorov hravo

nahradiť. Ako vy vidíte budúcnosť štúdia informatiky?

Závisí to od toho, akú budúcnosť očakávame. Keď si predstavíme, že AI dosiahne úroveň človeka, teda vytvoríme všeobecnú umelú inteligenciu (AGI), tak skoro všetku našu prácu bude vedieť urobiť. Zanikne tak veľmi veľa povolaní. Uplatnia sa možno len zabávači, umelci a speváci, lebo ich diela si možno budeme vážiť.

Podľa mňa je však pravdepodobnejšie, že sa mnohé zamestnania „iba“ veľmi zmenia. Vďaka AI nahradíme veľmi veľa mravčej roboty. Tej je veľmi veľa aj v programovaní. Keď chce človek niečo naprogramovať, uvedomí si, že len prekladá myšlienku do nejakého iného jazyka. To je práca, ktorú môžete dať urobiť junior developerovi.

S umelou inteligenciou teda napríklad potrebujeme oveľa menej začiatočných junior developerov. Preto sa juniori musia učiť byť oveľa strategickejší. Musia vedieť viac o softvérovej architektúre, viac o tom, ako celú aplikáciu zložiť a podobne. Budeme teda potrebovať viac ľudí, ktorí vedia strategickejšie rozmýšľať a ktorí vedia používať tieto nové nástroje.

Medzi piatimi knihami, ktoré ste vybrali, je aj knižka [*Artificial Intelligence: A Modern Approach*](#). Z nej ste sa učili už na Oxforde?

Áno. Profesori z nej vyberali jednotlivé kapitoly. Šprti ako ja si prečítali aj viac. Je to kniha, ktorá dá človeku veľmi dobré intro o tom, čo je umelá inteligencia, ako o nej rozmýšľať, kde sa končia ľudsky

písané algoritmy a kde sa začína umelá inteligencia, alebo kde sa to prelína so strojovým učením. Musí byť každá umelá inteligencia postavená na základe strojového učenia? Ako vyhodnocujeme, čo je inteligencia? A ako vyhodnocujeme, či je nejaký systém inteligentný? Všetky takéto otázky sú v knihe rozobraté.

Kniha vyšla prvýkrát už v 90. rokoch, teraz je k dispozícii už jej aktualizovaná štvrtá edícia. Tá kniha sa asi musela za ten čas úplne zmeniť, nie?

Základy, teória a definície budú stále veľmi podobné. Slovné spojenie umelá inteligencia vzniklo v 50. rokoch minulého storočia a veľa teoretických a filozofických problémov je dodnes rovnakých.

Je to vyše tisícstranová kniha. Nakoľko je zrozumiteľná aj pre človeka, ktorý neštudoval matematiku a informatiku na vysokej škole?

Veľká časť je napísaná naozaj veľmi prístupne. Pre človeka, ktorý nečíta odbornejšiu literatúru, to môže byť trošku suché. Prípadne môže mať dojem, že sa kniha príliš venuje presným definíciám. Ale inak je to prístupné.

Tomáš Halgaš. Foto – archív TH

V čom sa odlišuje fungovanie ľudskeho mozgu od fungovania umelej inteligencie?

Veľmi veľa nápadov aplikovaných do umelej inteligencie máme práve z prírody a z pochopenia toho, ako funguje mozog. Keď som bol na univerzite, tak jedni z najšikovnejších študentov umelej inteligencie mali bakalára z neurovedy. Prišli na to, že keď zmapujeme, ako funguje napríklad ucho, tak sa môžeme naučiť, ako môže umelá inteligencia lepšie spracovať zvuk. Veľa takýchto zistení sa dnes pridáva do umelej inteligencie.

Zaujímavé je aj, že ľudský mozog má stále o dosť väčšiu výkonnosť ako tie najväčšie počítače. Prebiehajú v ňom oveľa komplikovanejšie procesy. Každý neurón v mozgu je oveľa zložitejší ako umelý neurón, ktorý vieme naprogramovať. V počítači však vieme urobiť výpočty oveľa rýchlejšie.

O mozgu a umelej inteligencii je aj kniha [*How to Create a Mind*](#).

Na tejto knihe je krásne, že dá človeku skvelú intuíciu o tom, ako sa dajú neuróny poskladať a vytvoriť niečo oveľa inteligentnejšie. Mozog napríklad rozoznáva čiernu a bielu. Keď sa niekoľko neurónov spojí a vidia niečo čierne v rade a okolo toho biele, tak chápú, že je to čiara. A keď spojíme ďalšie neuróny, zistíme, že sú tam dve čiary. Keď sa pridajú ďalšie neuróny, zistíme, že je to písmeno L. Zrazu vidíme celé slovo a na vyššej úrovni viem, čo to slovo znamená. Tá kniha krásne vysvetľuje, čo všetko sa dá z malých súčiastok vyskladať.

Prečo je vývoj v oblasti umelej inteligencie taký dynamický? Čo bola tá kľúčová ingrediencia, vďaka ktorej dnes technológie tak rýchlo napredujú?

Keď som rozmýšľal, ktoré knihy odporučiť, tak som chcel vybrať aj *The Structure of Scientific Revolutions* ([*Struktúra vedeckých revolúcií*](#)) od Thomasa Kuhna. On opisuje momenty, keď musí nastať nejaký zlom, ktorý odomkne veľmi veľa nového vývoja. Potom sa to na nejaký čas zas zasekne a čakáme na novú inováciu. V podstate celá veda sa vyvíja takýmito skokmi.

Vidíme to veľmi pekne aj v umelej inteligencii. Na začiatku boli veľmi jednoduché štatistické algoritmy, veľa vecí to dokázalo urobiť, no potom sme sa zasekli. Neskôr sme vymysleli umelé neuróny, vďaka tomu sa podarilo veľa vecí a opäť sa to zaseklo. Potom prišli ešte možno jedna alebo dve podobné revolúcie.

Teraz nám veľa vecí odomkol vynález transformerov. Jeden revolučný vedecký článok tvrdil, že „attention is all you need“, teda pozornosť je všetko, čo treba. Takýto algoritmus je teda schopný zo všetkého, na čo sa pozerá, vybrať, na čo sa treba zamerať a z toho ponúkne nejaký výsledok.

Zistili sme, že upriamiť pozornosť je veľmi mocný mechanizmus, z ktorého vzniklo veľa algoritmov, napríklad rozpoznávanie objektov

v obrázkoch. A revolučná bola aplikácia tohto mechanizmu do prirodzeného jazyka. Pretože veľa z toho, ako ľudia rozmýšľajú, je zabudované do nášho jazyka. Takže vďaka tomu, že umelá inteligencia lepšie rozumie jazyku, vie napodobniť veľa ľudských schopností.

Tretia kniha, ktorú ste vybrali, má názov [The Alignment Problem](#). To v preklade znamená problém zosúladenia či zladenia. Jeho podstatou je, že chceme, aby AI napĺňala naše zamýšľané ciele. Ako sa to dá lepšie vysvetliť?

Tá knižka ukazuje, ako sa boríme s tým, aby sa umelá inteligencia naučila to, čo chceme, aby sa naučila. Problém totiž je, že sa učí veľa vecí, ktoré ani nechceme, aby vedela.

Kniha povedzme opisuje použitie umelej inteligencie v súdnictve. V USA už napríklad fungovali algoritmy, ktoré spracovali materiály o nejakom súdnom prípade. Počítač vypočítal, že riziko, že dotyčný daný čin spravil, je pomerne veľké, takže by sme ho mali dať do väzby.

Keď sa na to začali vedci pozeráť, zistili, že stačilo zmeniť rasu človeka. Teda, že ak boli ľudia napríklad afroamerického pôvodu, tak podľa algoritmu bolo pravdepodobnejšie, že čin spáchali. To sú veci, ktoré sa umelá inteligencia učí od ľudí, no nemala by sa to učiť. Tieto prípady totiž môžu mať štatistickú koreláciu, ale určite to nie je kauzalita.

Zaujímavé na tom je, že takéto veci by sa naučil aj človek. No my chceme, aby sa ich algoritmus naučil ešte lepšie ako človek. Teda, napríklad, aby odstránil naše vlastné predsudky.

Iný príklad je autodráha. V hre sa umiestnili kužele a auto malo jazdiť okolo nich. Avšak bola tam chyba, pretože skóre sa počítalo podľa

toho, koľkokrát auto prešlo posledným cieľovým kužeľom. Umelá inteligencia na to rýchlo prišla a začala sa točiť na mieste.

Prechádzala len koncovým kužeľom, pretože sa nepočítalo, či auto prešlo aj okolo všetkých ostatných kužeľov. Človeku je jasné, že toto sme nechceli, aby sa umelá inteligencia naučila.

Aké to môže mať dôsledky?

Najväčšie implikácie tohto alignment problému sa podľa mňa prejavili na sociálnych sieťach. Snažili sme sa vytvoriť umelú inteligenciu na odporúčanie príspevkov, aby sme mali úplne skvelú sociálnu sieť, na ktorej sa budeme spolu baviť. Ale ako to vysvetliť umelej inteligencii?

Keď sme chceli, aby umelá inteligencia maximalizovala množstvo aktivity, ktorú používatelia na sociálnych sieťach vykonávajú, veľmi rýchlo prišla na to, že sa to dá dosiahnuť tak, že bude odporúčať obsah, ktorý je šokujúci, negatívny či polarizujúci. Prvý veľký stret na svetovej úrovni s umelou inteligenciou sme teda vôbec nezvládli.

Sme lepšie pripravení na ďalšie strety?

Keď hovorím o rizikách umelej inteligencie, väčšinou hovorím o troch úrovniach. Prvou je práve alignment problem, pretože je veľmi ťažké vysvetliť stroju komplexitu ľudského života. Človek je plný rozporov. Na jednej strane máme napríklad ambície, na druhej strane nás baví sadnúť si pred televízor a pozerať hokej. Keď sa na to pozrie umelá inteligencia, tak to nedáva zmysel. To sú veci, ktoré sa ťažko opisujú, ale robia nás ľudskými.

Druhá úroveň rizík je extrémna automatizácia. Pomocou strojov sa nám podarilo zautomatizovať veľa ľudskej práce alebo sa nám podarilo zautomatizovať niektoré procesy na úradoch a stačí si ich ľahko vyťukať na internete. Teraz však hrozí, že zautomatizujeme nejaké veci na úkor veľmi veľa pracovných miest. Prichádza to v bezprecedentnej miere. V USA sa viac ako 60 percent ľudí bojí o svoju robotu pre umelú inteligenciu. Na Slovensku? Tu sa o tom ešte len začína rozprávať. Ľudia sú na to extrémne nepripravení.

A čo je tretia úroveň rizík?

Nechcem strašiť čitateľov, ale veľa veľmi bystrých ľudí verí tomu, že nám hrozí existenčné riziko. Vo veľmi dramatickom poňatí by to mohlo vyzerať ako *Terminátor* alebo ako film *Transcendence*, ktorý by som odporúčal si pozrieť. Ukazuje totiž, ako môže umelá inteligencia veľmi rýchlo zrýchliť vedecký pokrok, ako rýchlo sa začne zapájať do politiky a ekonomiky, a že ľahko ovládne mnohé ľudské inštitúcie. Myslím si, že je šanca, že môžeme čeliť aj existenčnému riziku. Umelá inteligencia je totiž extrémne mocný nástroj.

Štvrtá kniha, ktorú odporúčate, sa volá [Život 3.0](#). O čom je?

Je to naozaj dobrá štúdia asi všetkých možností, ktoré nás môžu čakať. Problémom je, že veľmi veľa scenárov je dystopických. Nepáči sa mi napríklad budúcnosť, v ktorej už človek nemá ako pridať nejakú hodnotu a len si užíva. Nepáči sa mi ani budúcnosť, v ktorej oveľa inteligentnejšia umelá inteligencia ovládne ľudstvo a budeme žiť v nejakom Matrixe.

Ukazuje sa, že na scenároch, ktoré by sa nám mohli páčiť, treba veľmi pracovať, alebo sa treba aspoň veľmi aktívne usilovať o to, aby niekto umelú inteligenciu nezneužil.

Veríte tomu, že sa nám podarí dosiahnuť tie pozitívnejšie scenáre?

Som ľahko až mierne optimistický. Veľa šikovných a chytrých ľudí sa

týmto problémom venuje. Poznám osobne ľudí aj vo firmách, ako je OpenAI, ktorým na týchto veciach záleží a aktívne sa venujú AI bezpečnosti. Existuje teda pomerne veľká nádej, že aj pomocou týchto veľkých firiem môžeme umelú inteligenciu kontrolovať.

Spomenul by som ešte knižku – [The Sovereign Individual](#), ktorá hovorí o tom, že čím viac máme možností, ako internet, bitcoin, tým sme menej závislí od štátu. Aj preto vraj budú štáty stále slabšie a slabšie. Oveľa viac moci bude v rukách firiem a jednotlivcov. To znamená, že sa budeme musieť spoľahnúť na to, že pár veľmi mocných ľudí nás bude hnať pozitívnym smerom.

Posledná, piata knižka, ktorú ste vybrali, je [Spravedlnost](#). Nie je to kniha o umelej inteligencii, prečo ste ju teda vybrali?

Je to knižka o ľudských hodnotách a o tom, čo je spravodlivosť. Opisuje, ako sa v histórii vyvíjalo, čo sme považovali za spravodlivé. A ponúka veľa otázok na zamyslenie. Človek si pri jej čítaní uvedomí, že ani my ľudia sa často nevieme dohodnúť, čo je spravodlivosť, a aké by mali byť naše hodnoty. Ako teda chceme dať nejaké hodnoty umelej inteligencii?

Aké východisko vidíte?

Bude to na týchto pár veľmi vplyvných jednotlivcoch, aby dali dokopy nejaké riešenie.

Očakávate, že by tak mohli vzniknúť nejaké globálne štandardy, na ktorých by sa dohodol celý svet?

Skôr budeme vidieť nejaké západné modely, ktoré budú reflektovať západné hodnoty a kultúru. A nejaké východné modely, ktoré budú nastavené úplne inak.

Ľudia vás možno poznajú ako Slováka, ktorý predal svoju firmu Twitteru. A potom možno z tlačovej besedy Mikuláša Dzurindu, keď oznamoval návrat do politiky a vás predstavil ako jeden z talentov v jeho strane. Ako dnes vidíte svoju politickú budúcnosť?

Som tomu otvorený, ale je to otázka možností a načasovania. Je veľa vecí, ktoré potrebujeme urobiť, ale najprv musíme vyriešiť niektoré oveľa závažnejšie politické otázky, v ktorých nemám až takú expertízu. Napríklad, aké má byť smerovanie krajiny. Keď ich vyriešime, možno sa vytvorí priestor aj pre človeka, ako som ja, ktorý by chcel pomôcť pripraviť Slovensko na umelú inteligenciu a podobne.

Tomáš Halgaš (1993) vyštudoval matematiku a informatiku na univerzite v Oxforde. Krátko pracoval pre Facebook. Spoluzakladal spoločnosť Sphere, ktorú v roku 2021 [kúpil](#) Twitter. Dnes sa venuje svojej ďalšej firme Sutro, ktorá pomocou umelej inteligencie vyvíja aplikácie.

Rubrika *Päť kníh* je inšpirovaná britským projektom FiveBooks.com a vychádza so súhlasom autorov tohto projektu. Nové vydanie publikujeme vždy v stredu. Ak si kúpite knihu cez niektorý z odkazov v tomto texte, Denník N získa za nákup malú províziu. Ďakujeme, že podporujete spisovateľov, kníhkupcov aj Denník N.