

Kňaz venujúci sa téme AI: Umelá inteligencia pôsobí na ľudské slabosti

S rímskokatolíckym kňazom a zároveň informatikom Petrom Šantavým sme sa rozprávali na tému umelej inteligencie.

Peter Šantavý je kňazom v bratislavskom Kostole Nanebovzatia Panny Márie. Okrem toho vedie Centrum informačných technológií Bratislavskej arcidiecézy, má na starosti informačné technológie Rímskokatolíckej cyrilometodskej bohosloveckej fakulty UK a participuje na riešení kybernetickej bezpečnosti Univerzity Komenského. Okrem iného sa venuje téme bezpečnosti a etiky umelej inteligencie (AI). O nej sa v tomto rozhovore budeme rozprávať.

Boli ste najprv kňazom, ktorý sa začal zaoberať IT, alebo sa z ITčkára stal kňaz?

Vyštudoval som najprv elektrotechnickú fakultu, pracoval ako softvérový vývojár a až neskôr som vstúpil do seminára.

Sama som v živote stretla viacerých ľudí pracujúcich v IT, ktorým pri tej vysoko technickej práci chýbalo nejaké duchovno. Bol to aj váš prípad?

Ani nie i keď treba povedať, že na elektrotechnickej fakulte a „matfyz“ bolo mnoho ľudí, ktorí nejakým spôsobom inklinovali či už k niečomu duchovnému, alebo k psychológii. Pred osemdesiatymdeviatym sa dokonca o matematicko-fyzikálnej fakulte a o elektro fakulte hovorilo, že sú to také bašty klerikalizmu. U mňa však všetko začalo až po vysokej škole. Išlo o postupné kroky, ktoré viedli až k semináru.

Kňaz, ktorý je zároveň expertom na informačné technológie, je v našich podmienkach vcelku neobvyklá kombinácia.

Je to taká zaujímavá kombinácia, možno niekedy neľahká, lebo človek sedí na viacerých stoličkách. Tieto úlohy sa však dajú sklbiť, pretože mám na starosti malý kostolík. Správa veľkej farnosti by s mojou druhou prácou nešla veľmi dokopy. V poslednom období zároveň chodievam do rôznych kresťanských organizácií robiť prednášky o umelej inteligencii. V rámci AI sa okrem etiky a bezpečnosti osobitne zaujímam aj o tému všeobecnej umelej inteligencie (AGI).

Keď si AGI zadáte do vyhľadávača, dozviete sa, že ide o hypotetickú formu AI, ktorá by mohla dosiahnuť rovnakú alebo vyššiu intelligenčnú úroveň ako človek. Vy si myslíte, že raz ľudstvo takúto AI vyvinie?

Na túto tému sú viaceré pohľady. Mnohí po vzniku AGI túžia a zároveň predpovedajú, že sa ju jedného dňa podarí vytvoriť. Sú i pohľady skeptické, ktoré hovoria o tom, že takú AI, ktorá by naplno nahradila človeka, nevyvinieme. Ja patríam medzi skeptikov. Ľudia totiž nie sú akési biologické stroje, ľudská inteligencia je komplexná a prepojená s ľudskou dušou. Môžeme vyvinúť akurát stroje, ktoré budú ľudskú inteligenciu simulovať. Alebo budú mať istú formu inteligencie, lenže úplne inú, než majú ľudia, na inej úrovni. Samozrejme, aj takáto strojová inteligencia, ktorá nedosiahne plnosť tej ľudskej, môže byť veľmi nebezpečná.

Keď už ste načali tému nebezpečenstva, mnohí ľudia sa boja, čo všetko by umelá inteligencia mohla v budúcnosti spôsobiť. Ide podľa vás skôr o neporozumenie téme, keďže samotné fungovanie AI je pre väčšinu ľudí nezrozumiteľné, alebo nám naozaj hrozia katastrofické scenáre?

Čo sa týka rizík, s obľubou na svojich prednáškach ukazujem jeden slajd, na ktorom sú znázornené dva prieniky ľudskej inteligencie a umelej inteligencie. Jeden je “vysoký”, ten by nastal vtedy, keby umelá superinteligencia presiahla schopnosti človeka. Druhý je “nízky” prienik. V rámci neho umelá inteligencia pôsobí na ľudské slabosti a premáha tak človeka.

Tento princíp ste spomínali aj vo svojej knihe. Rozprávame sa o umelej inteligencii, ktorá sa už, napríklad vo forme algoritmov, používa. Narážame na ňu v každodennom živote a ani o tom mnohokrát nevieme, napríklad pri používaní sociálnych sietí, ktoré nám s jej pomocou zobrazujú obsah. Aké tu vidíte nebezpečenstvá?

Tento druh umelej inteligencie zbiera dáta z reálneho sveta a človeka monitoruje, oberá o súkromie a vnútornú slobodu, predikuje naše správanie, dokáže nás aj k určitému správaniu usmerňovať. Vytvára tiež sociálne bubliny, v dôsledku ktorých dochádza napríklad k relativizácii pravdy. Osobitným problémom je závislosť, strata kognitívnych schopností (napríklad digitálna demencia či kognitívna kapitulácia) alebo uzavretie sa do virtuálnych svetov, a to najmä u mladých ľudí, ktorých mozgy sú ešte “plastické” a psychika krehká.

Znie mi to akoby sa mnohí desili akejsi dystopickej AI budúcnosti, pričom si neuvedomujú, že ju v istom zmysle dávno žijeme. Umelá inteligencia nám nielen servíruje obsah na sociálnych sieťach, mnohí svoje problémy zverujú AI konzultantom, ako je ChatGPT. Čo s tým?

Treba zdôrazňovať, že generatívne systémy, akým je aj ChatGPT, iba simulujú myslenie. Tieto modely aj zlyhávajú iným spôsobom, než ľudia, nerozlišujú dobré od zlého a neuvedomujú si dôsledky. Nevedia tiež povedať “neviem”, a namiesto toho “halucinujú”, teda si vymýšľajú, prípadne môžu ponúkať neobjektívne, čiastočné či nesprávne odpovede. Netreba ani zabúdať na to, že tieto služby uchovávajú dáta, ktoré im človek zverí. Vráťane tých najintímnejších. Stále zároveň ide o stroje, ktoré musia byť vždy pod kontrolou človeka, ktorý ich vie správne používať a výstupy preveruje.

Mám dojem, že práve skutočnosť, že ide o stroje, si niektorí ľudia prestali uvedomovať. Je to aj pochopiteľné. Keď sa niečo správa ako človek, a v prípade AI spoločníkov to aj vyzerá a znie ako človek, je normálne pripisovať takejto simulácii ľudské vlastnosti.

Táto virtualizácia vzťahov je podľa mňa v súčasnosti silným fenoménom. Pôvodne som si myslel, že pôjde len o niečo okrajové. Pre ľudí, ktorí nejakým spôsobom potrebujú utiecť z tohto reálneho sveta. Ale medzi mladými si všímam, že mnohí vzťahy s AI spoločníkmi najprv skúšajú ako takú zábavku. Potom zistia, že takéto vzťahy sú oveľa jednoduchšie než tie ľudské. A tak si aj človek, ktorý by k podobným únikom normálne nemal sklon, môže s týmito systémami vytvoriť toxický vzťah.

Ako z toho von?

Aby sme sa tomu vyhli, je veľmi dôležitá digitálna askéza. Musíme tiež ľudí vtiahnuť do reálneho života. Lebo keď budeme skupine mladých o živote v realite iba hovoriť a oni pritom budú pozerat' do mobilu, nebude to fungovať. No ak ich vyťahujeme do reálneho sveta, napríklad formou zážitkov alebo tvorivej činnosti, aspoň čiastočne by im to mohlo pomôcť. Jednoducho by mali vo virtuálnom svete tráviť obmedzené množstvo času, budovať digitálnu askézu.

Tento problém sa však stále viac týka aj dospelých. Čo by mohlo podľa vás pomôcť ako osveta pre ľudí, ktorí sa stále viac spoliehajú napríklad na rady ChatuGPT a podobných služieb?

Neustále musíme opakovať, že AI iba simuluje myslenie. Pôsobí to, akoby stroj rozumel, chápal, ponúkal premyslené odpovede a nemohol sa myliť. No nie je to tak. Niektoré služby AI konzultantov už obsahujú podobné upozornenia. Ale podľa mňa sa celý tento komplex nezmení dovedy, kým jeho základným cieľom bude monetizácia. Chýbajú tu zároveň jasné etické pravidlá, celá oblasť AI je ako Divoký západ. Sama sa táto oblasť nezreguluje.

Vo svojej knihe ste spomínali nariadenie Európskej únie o umelej inteligencii. Je podľa vás dobrým riešením?

V čase, keď som svoju knihu písal, teda v roku 2022, sa mi takým javil. Postupne sa však ukazuje, že to nariadenie nie je úplne univerzálne. Regulácia zároveň musí byť nastavená tak, aby na jednej strane chránila základné etické pravidlá a normy, na strane druhej aby umožňovala rozvoj.

Takže máte dojem, že to EÚ preregulovala?

V rámci niektorých aspektov áno. Niečo je tam tiež ťažko vymožiteľné. Popri dobrých ustanoveniach sa tam nachádza aj viacero diskutabilných. V oblasti AI zároveň prebieha taký búrlivý vývoj, že podobné regulácie už čoskoro nemusia stačiť.

Ďalšou snahou o nastavenie etických rámcov je najnovšia encyklika pápeža Leva XIV. *Magnifica Humanitas*. Čo na ňu hovoríte?

Ide o veľmi kvalitný dokument, ktorý podľa mňa v rámci digitálnej revolúcie bude podobne prelomový ako *Rerum Novarum* Leva XIII. v kontexte prvej priemyselnej revolúcie v devätnástom storočí. Medzi oboma revolúciami je rozdiel, lebo doterajšie priemyselné revolúcie znamenali skokovú zmenu a tým pádom aj skokovú adaptáciu spoločnosti a civilizácie. Teraz sme skôr v stave permanentnej vedecko-technologickej revolúcie. Naša civilizácia sa musí adaptovať neustále. Kým si na niečo zvykneme, už sme konfrontovaní s ďalšou zmenou.

V čom vidíte najväčší prínos encykliky?

Tento dokument veľmi pekne načrtol etický rámec a zdôraznil hodnotu ľudského bytia. Doposiaľ sme boli svedkami skôr čiastočných riešení, kdežto encyklika priniesla skutočne komplexnú odpoveď. Je tam antropologický akcent v kontexte Božieho stvoriteľského diela, Božej milosti a Božieho pozvania k svätosti, aby sme vnímali jedinečnosť človeka, ľudského bytia, ktorú nemôže presiahnuť nejaký technologický rozvoj, transhumanizmus alebo strojová inteligencia.

Niektorí kritici naznačujú, že cirkev je a priori proti technologickému rozvoju. Je to tak?

Nie. Napríklad encyklika vôbec nie je proti technológiám. Zdôrazňuje, aby sme ich rozvíjali a využívali na dobro. Aby išlo o niečo, čo slúži človeku a nie niečo, čo ľudstvo ohrozuje alebo rozdeľuje.

Keď hovoríme o tej službe človeku, v čom vidíte najväčší prínos umelej inteligencie pre ľudstvo?

V akejkoľvek oblasti, ktorú by sme začali rozoberať, nájdeme veľmi dobré aplikácie. Čím viac sa tieto technológie rozvíjajú, tým viac jestvuje aplikácií, ktoré podporujú človeka, teda napríklad vzdelávanie, zdravie, ochranu, alebo vedecký výskum. A prakticky v každej z oblastí, ktoré by sme spomenuli, nájdeme hneď aj veľké riziká.

Spomeniete si aj na nejaké konkrétne príklady využitia AI na dobré veci?

Ja osobitne veľmi obdivujem dosiahnuté úspechy nasadenia umelej inteligencie vo vedeckom výskume. Napríklad v astrofyzike je to úplne geniálne. Podarilo sa spracovať množstvo dát z rôznych zdrojov, či už satelitov alebo teleskopov, ktoré dovtedy nebolo možné spracovať. Ďalej by som spomenul oblasť medicíny, hoci tu sa mi hneď natískajú aj riziká zneužitia. Ale aj v sociálnej starostlivosti má AI veľkú perspektívu. Napríklad v rámci pomoci ľuďom s Alzheimerovou chorobou. Toto je, mimochodom, z mojej strany asi jediná oblasť, kde akceptujem humanoidné systémy umelej inteligencie.

Nebezpečenstvá AI sme už načrtli vyššie, ale keď sa rozprávame o jej prínosoch, nedá mi nezablúdiť aj k rizikám. Ktoré sú podľa vás tie najväčšie?

Vidím štyri veľké oblasti problémov AI, pred ktorými treba ľudstvo chrániť. Prvou je ochrana súkromia, druhou zápas o slobodu, teda ochrana človeka pred predikciou a manipuláciou, treťou

ochrana kognitívnych schopností človeka pred digitálnou demenciou, kognitívnou kapituláciou (strata kritického myslenia a neschopnosť robiť rozhodnutia), mozgovou hnilobou (brain rot) a digitálnym rozdelením. A štvrtou oblasťou sú virtuálne vzťahy a riziko straty zmyslu pre realitu. Ďalšou obrovskou hrozbou je nasadenie AI vo vojne, respektíve v armáde. Vzniká tu napríklad riziko vzniku takzvanej hypervojny, teda vojny riadenej algoritmiami umelej inteligencie. Ľudské rozhodovanie v tomto druhu konflikte zaniká, čo je veľmi nebezpečné.

Koniec ľudského rozhodovania však hrozí aj v ďalších oblastiach, pokiaľ sa budeme čoraz viac spoliehať na AI.

Jedným z katastrofických scenárov vývoja AI je práve ten, kedy sa ľudstvo tak spoľahne na umelú inteligenciu, že sa na ňu kognitívne “zavesíme”, a tým pádom prebehne kognitívna aj vzťahová degradácia. Tá by mohla radikálnym spôsobom napredovať a mať za následok morfológické zmeny na mozgu vrátane tých medzigeneračných, ktoré by boli ťažko vratné.

Čo môže každý z nás spraviť pre to, aby tento scenár nenastal?

Mali by sme vypočuť pozvanie reálne žiť. Vo svete skutočných tvárí a hlasov. A praktizovať digitálnu askézu.

Keďže pracujete s počítačovými systémami, digitálna askéza môže byť problematická. Ako to riešite?

Snažím sa vždy, keď je to možné, si vyhradiť čas bez týchto technológií. Okrem toho trénujem pamäť, aby som sa nespoliehal na techniku. Už len taká maličkosť, že keď si idem vyzdvihnúť zásielku, snažím sa zapamätať jej číselný kód a nie loviť ho v mobile. Zdá sa to byť málo, ale aj takto je možné prispievať k udržiavaniu kognitívnych schopností.

Na záver ešte otázka: Ako môže podľa vás byť cirkev nápomocná v oblasti etiky AI?

Môže ľudom pomáhať vymaniť sa z technokratickej totality a s ňou spojenými rizikami. Cirkev by tiež mala byť hlasom prezentujúcim Božie zjavenie a Božie pozvanie, aby sme chránili ľudskú dôstojnosť a pozývali človeka kráčať cestou vykúpenia. Mali by sme budovať človečenstvo, ktoré sa dvíha nad hriech a nad všetko to, čo je okolo nás, aj vďaka Božej milosti a osobnej svätosti. A potom, samozrejme, cirkev môže prispieť aj šírením osvetu a vzdelávania v oblasti etiky využívania AI. V charitatívnej oblasti pracujeme na zmiernení digitálneho rozdelenia, aby neboli tí, ktorí sú marginalizovaní či chudobní, na základe svojho technologického deficitu využívania, zneužívaní či vykorisťovaní. A napokon oblasť vojenského využitia AI – ako to badať na konfliktoch na Ukrajine, v Gaze a v Iráne – je veľkým pozvaním pracovať so všetkými ľuďmi dobrej vôle na mierovom a etickom využívaní systémov AI.

Myslím si, že na základe aktivít, ktoré cirkev v posledných rokoch mala (*Rome call for AI ethics, Antiqua et nova, Magnifica humanitas,...*), môže sa zapájať aj do aktivít, ktoré povedú k reguláciám návrhu, tvorby, nasadenia a využívania systémov AI na regionálnej i celosvetovej úrovni.