Jan Stuchlý z Kliniky dětské hematologie a onkologie

Možnosti nástrojů umělé inteligence jsou již delší dobu využívány v medicíně, tedy i na lékařských fakultách UK. O tom, jak AI mění současnou vědu, proč je důležitější správně položit otázku než znát odpověď a jak mohou podobné nástroje pomoci při zkoumání leukemií, jsme hovořili s Janem Stuchlým z Kliniky dětské hematologie a onkologie 2. LF UK a FN Motol a Homolka.

Ve vaší práci se pojí lékařství, věda, technologie i matematika. Kdy jste začal v těchto oblastech pracovat s umělou inteligencí a byla nějaká konkrétní výzva, která vás k tomu přivedla?

Vlastně jsem s tím začal pracovat ještě dřív, než se o tom běžně mluvilo jako o umělé inteligenci. Pracovali jsme s relativně jednoduchými neuronovými sítěmi.
Když jsem začínal s něčím, co bychom dnes mohli označit za AI, bylo to v souvislosti s vizualizací a analýzou velmi komplexních biologických dat.

V biologii bývá často problém už jen uložit a zpracovat objem dat, který vzniká během analýz. Z několika pacientů mohou vzniknout terabajty dat, která není jednoduché nikam nahrát ani ekonomicky provozovat v externích službách.

Právě proto mi dnes připadá nejzajímavější spojení velkých jazykových modelů s lokální výpočetní infrastrukturou. Máme výkonné výpočetní uzly, velká úložiště a k tomu jazykový model, který dokáže pomáhat s programováním, řízením analýz i obsluhou celého procesu. Pro mě je nejzajímavější právě potenciál těchto systémů při práci s reálnými vědeckými daty a výpočetními zdroji.

A používáte nástroje umělé inteligence běžně při práci nebo jen někdy na konkrétní, specifické úkoly?

Používám ji v podstatě každý den a poměrně rutinně, zejména na programování. Dnes už sám programuju hlavně tehdy, když si potřebuji napsat nějakou jednoduchou funkci nebo když mě zajímá konkrétní algoritmický problém.
Umělou inteligenci používám i na věci, u kterých mě vlastně překvapilo, jak dobře funguje. Když se učím novou matematickou metodu z učebnic nebo odborných článků, bývají jejich součástí cvičení, která si člověk potřebuje projít, aby látce skutečně porozuměl. Někdy se přitom může zaseknout na jediném příkladu klidně na celé odpoledne.

Ve velké většině případů dnes dokážou velké jazykové modely takové úlohy vyřešit správně a pomáhají mi pochopit postup výrazně rychleji. A to ještě před rokem na mě velké jazykové modely v podobných situacích působily trochu jako student, který se snaží zakrýt, že látce úplně nerozumí. Dnes mám mnohem častěji pocit, že skutečně něco počítají a řeší.

A je naopak něco, co vás překvapuje, že umělá inteligence stále neumí?

Nejsem tím úplně překvapený, ale myslím si, že umělá inteligence zatím ani nevykročila směrem k tomu, aby přicházela s vlastními nápady. Nedávno jsem četl vzpomínku na matematika Stanisława Ulama. Jednoho z jeho kolegů se ptali, proč byl tak výjimečný, a on odpověděl: „Protože měl ty nejkrásnější otázky.“
Velké jazykové modely jsou dnes stále lepší v odpovídání. Když člověk položí dostatečně jednoznačnou otázku, dokážou pomoci s překvapivě širokým spektrem problémů. Skutečný pokrok ale často nespočívá v tom, že vyřešíte dobře definovanou otázku. Důležité je položit si tu správnou otázku. A právě v tom zatím vidím jejich limit.

Váháte někdy kvůli těm limitacím nad použitím nástrojů umělé inteligence? Jsou věci, které ji zatím radši nesvěříte?

Asi bych řekl, že když člověk studuje matematiku, často se učí právě tím, že dlouho přemýšlí nad cvičením a hledá vlastní řešení, láme si s tím hlavu. Pokud si všechno nechá okamžitě vysvětlit, může o část tohoto procesu přijít. Je to podobné, jako kdyby si nechal poradit od chytřejšího spolužáka.
Na druhou stranu existují úlohy, nad kterými lze strávit týden a stejně se nikam nedostat. Pak je otázka, jestli není lepší nechat si poradit. Najít správnou rovnováhu mezi pomocí, která člověka posouvá dál, a pomocí, která ho zbavuje přemýšlení, je podle mě jedna z nejtěžších věcí při používání AI a často nad tím sám přemýšlím.

Působíte na druhé lékařské fakultě Univerzity Karlovy. Máte nějakou představu třeba o tom, jak AI v blízké budoucnosti promění vysoké školy?

Myslím, že s nástupem umělé inteligence budeme muset znovu promyslet, co vlastně považujeme za inteligenci nebo vzdělanost. V určité době byl vzdělaný člověk ten, kdo přečetl hodně knih, uměl vyhledávat informace v knihovnách nebo třeba skvěle počítal. Pak přišly kalkulačky, počítače a internet a význam těchto dovedností se postupně proměnil. Předpokládám, že podobná změna nás čeká i teď.
Samotné množství znalostí nebo zvládnutých technik už nebude tak významnou výhodou jako dřív.

Claude Science na univerzitním clusteru

V RSE se věnujete mimo jiné i zpřístupňování nástrojů umělé inteligence akademikům a výzkumníkům. Na univerzitním clusteru je nyní dostupný také Claude Science. Jak si mám tento nástroj jako laik představit?

Claude Science je příkladem takového agentního systému pro vědeckou práci. Nejde jen o jazykový model, který odpovídá na otázky, ale o nástroj, který dokáže koordinovat práci s dalšími specializovanými aplikacemi a analytickými nástroji.

Výzkumník může jednoduše popsat, co chce udělat, a systém navrhne vhodný postup, vybere nástroje, připraví potřebné skripty a podobně. Výrazně se tím snižuje ta vstupní bariéra při práci se složitějšími výzkumnými nástroji.
Na univerzitním clusteru je dnes Claude Science dostupný v testovacím režimu a akademici i výzkumníci si mohou jeho možnosti sami vyzkoušet.

Claude Science na univerzitním clusteru

Tým Research Software Engineering Univerzity Karlovy (RSE), se kterým spolupracuje Jan Stuchlý, nově zpřístupnil na univerzitním výpočetním clusteru nástroj Claude Science, specializovaný AI systém určený pro vědeckou práci.

Na rozdíl od běžných chatbotů dokáže spolupracovat s analytickými a výzkumnými nástroji, připravovat skripty nebo navrhovat postupy při zpracování dat. Akademici a výzkumníci UK si jej nyní mohou v testovacím režimu vyzkoušet přímo v univerzitní infrastruktuře, bez nutnosti využívat externí komerční služby.

Znamená to, že už je využití AI prakticky vyřešené?

Z pohledu samotných modelů jsme se posunuli dost daleko. Hlavní výzvou, kterou teď řešíme, ale nejsou ani tak schopnosti AI jako spíš funkční bezpečné prostředí, ve kterém ji chceme využívat.

Ve výzkumu často pracujeme s citlivými daty, například s údaji o pacientech. V takovém případě není vždy možné jednoduše využít běžné komerční AI služby. Potřebujeme mít jistotu, že práce s daty odpovídá legislativním požadavkům, nárokům na GDPR a že celý proces je bezpečný. Jednoduše mít k dispozici model, ke kterému bych mohl svá data připojit způsobem, u něhož mám jistotu, že splňuje všechny požadavky na ochranu dat.

Jednou z cest je provozovat modely lokálně, což je ale velmi náročné na hardware. Proto nyní zkoumáme také možnosti využití evropských cloudových služeb, kde je garantováno, že data podléhají evropské legislativě.

To mi trochu zní i jako propojení technologických otázek a administrativy, řešení různých pravidel...

To je právě ono, nejde jen o samotné modely, ale celou infrastrukturu kolem nich. Vedle technického napojení je potřeba zajistit správu přístupů, bezpečnost, kontrolu nákladů i propojení s univerzitními systémy.
Naším cílem je zpřístupnit výzkumníkům moderní AI nástroje tak, aby je mohli využívat bezpečně a bez nutnosti řešit technické nebo legislativní obtíže.

AI a výzkum liniové plasticity leukémie

Ve svém vlastním výzkumu se zabýváte takzvanou liniovou plasticitou leukemií. Co si pod tím lze představit?

Zjednodušeně řečeno se snažíme pochopit, jak se leukemické buňky mění v průběhu nemoci a léčby. Zajímá nás, proč jsou některé z nich schopné měnit svou identitu a jak přesně tento proces probíhá.
U některých typů leukemií totiž pozorujeme, že buňky, které původně připomínají jeden typ buněk imunitního systému, získávají vlastnosti jiného typu. Díky moderním metodám dnes dokážeme sledovat jednotlivé buňky a pozorovat, jak se během této proměny mění jejich genetický program.
Je to biologická otázka, která se ale vlastně stává i otázkou matematickou a výpočetní. Pracujeme s obrovským množstvím dat a snažíme se z nich rekonstruovat cestu, po které se buňky během své proměny ubírají. Právě proto se asi přirozeně nabízí využití pokročilých metod umělé inteligence.

Mohl byste to využití umělá inteligence při zkoumání leukemických buněk přiblížit?

Na jedné úrovni pomáhá se samotným zpracováním dat. Když například změříme single-cell data, výsledkem nejsou žádné přehledné grafy nebo tabulky. Z laboratoře dostaneme velké textové soubory obsahující miliony krátkých

genetických sekvencí. U každé sekvence víme, ze které buňky pochází, ale samotná data jsou pro člověka prakticky nečitelná.

V tomto okamžiku začíná bioinformatická analýza. Prvním krokem bývá zjistit, ke kterým genům jednotlivé sekvence patří. K tomu existují specializované nástroje, které sekvence porovnají s referenčním lidským genomem a určí jejich původ.
Jazykový model tyto úlohy sám neřeší. Funguje spíše jako inteligentní rozhraní nad existujícími bioinformatickými metodami. Rozpozná, jaký nástroj je pro daný krok vhodný, spustí jej a následně pracuje s výsledky.

Takže AI nenahrazuje odborné nástroje, ale spíše usnadňuje práci s nimi?

Je to tak. Ve skutečnosti výzkumník nepotřebuje jen samotný jazykový model. Potřebuje celou řadu bioinformatických nástrojů, které analýzu provádějí. Před nástupem jazykových modelů člověk většinou pracoval přímo v terminálu, zadával příkazy nebo si připravoval vlastní pipeline. U single-cell dat je navíc obtížné vytvořit univerzální postup, který by fungoval stejně pro každý experiment, takže tuto práci často musel dělat zkušený bioinformatik.
Dnes mohou jazykové modely fungovat jako prostředník mezi výzkumníkem a těmito nástroji. Dokážou připravit potřebné skripty, spustit jednotlivé kroky analýzy nebo pomoci s instalací nových metod. Výzkumník se tak může více soustředit na samotnou biologickou otázku a méně na technické detaily.

Kde vidíte největší potenciál nástrojů AI z pohledu samotné vědy?

Nejzajímavější je pro mě možnost využít je přímo při hledání odpovědí na biologické otázky. Když se například snažíme zjistit, jakou cestou se leukemické buňky během své proměny ubírají, existují desítky různých matematických modelů, které se na problém dívají z různých úhlů.
Dnes je velmi obtížné všechny tyto přístupy systematicky porovnávat. Do budoucna si ale umím představit, že jazykové modely budou schopné samostatně vyhledávat vhodné metody, aplikovat je na konkrétní data a pomáhat výzkumníkům hledat ty nejlépe fungující. To by mohlo výrazně urychlit cestu od dat k novým biologickým poznatkům.

Když se podíváte za dnešní možnosti AI, co považujete za opravdu zajímavý směr jejího dalšího vývoje?

Dnes už jazykové modely dokážou pracovat s existujícími nástroji a využívat je při analýze dat. Mě ale začíná zajímat otázka, jestli se mohou posunout dál a vytvářet vlastní matematické modely.

Řada biologických problémů se nakonec převádí do matematické podoby. Už dnes vidíme, že jazykové modely dokážou navrhovat hypotézy nebo nové analytické postupy. Nemyslím si, že by to zatím fungovalo rutinně, ale právě tady podle mě leží jedna z nejzajímavějších oblastí současného vývoje.

Představte si, že by model dokázal navrhovat metody pro analýzu dat a využívat přitom matematické principy z oborů, na které by člověk třeba vůbec nenarazil. Stále častěji se ukazuje, že řešení biologických problémů mohou přicházet z úplně jiných oblastí matematiky nebo fyziky. Tahle představa mě poslední dobou jako výzkumníka fascinuje. Tak uvidíme, kam se to bude posouvat.

Mgr. Jan Stuchlý, Ph.D.

Mgr. Jan Stuchlý, Ph.D. působí na Klinice dětské hematologie a onkologie 2. lékařské fakulty Univerzity Karlovy a Fakultní nemocnice Motol a Homolka a zároveň v týmu Research Software Engineering.

Na pomezí bioinformatiky, matematiky a výpočetních věd pomáhá budovat infrastrukturu, která umožňuje vědcům bezpečně využívat moderní nástroje umělé inteligence při výzkumu. Zároveň se věnuje vlastnímu výzkumu leukemií, konkrétně studiu takzvané liniové plasticity, tedy proměn identity leukemických buněk během nemoci a léčby.