Tady je Matěj Skalický a tohle je Vinohradská 12.
Tisíc expertů včetně Ilona Maska se psal opetici pro pozastavění vývoje umělých inteligenci.
A to, že popeš na sobě tuhle bundu nikdy neměl a fotografie je dílem umělej inteligence.
Tady je se stala první západní zemí, kde byla zablokovaná nejznámější internetová četovací služba ČED GPT.
Popeš v bílé pérovce, kůz kódů pro programátory, překlat textu a odpověď snad na všechno.
Je uměla inteligence atomovou bombou 21. století, proč je potřeba jí lychotit a co se na ní nezdá Ilonu Maskovy.
Ptám se kolegy datového novináře českého rozhlasu Michala Kašpárka.
Dnes je úterí čtvrtého dubna.
Ahoj Michale, vítej ve Vinohradské 12.
Ahoj a dobrý den všem, kteří posluchají.
Vymknul se vývoj umělé inteligence kontrole?
Nemyslím si, že by se už vymknul a spodle toho, co víme, na základě těch modelů, kterých my můžeme pracovat.
Ale myslím si, že ta otázka je na jednom mnohem víc na místě, než byla ještě před pár lety.
Já jsem četl takový komentář v brickém deníku The Telegraph od historika Dana Snowa,
že je to teď stejně nebezpečné jako projekt Manhattan,
tedy někde išli americký utajený vývoj nukleární bomby,
že stejně jako nukleární revoluce může podle něj uměla inteligence změnit svět,
jak ho známe, jen zatím nevíme jak, tak není to podle tebe trochu přitažené za vlasy.
Myslím, že není, myslím, že je to docela dobré připodobnění.
Projek Manhattan obnášel jen takovou situaci, kdy část větřů, která vyvělá tomovou bombu,
si nebyla úplnějstá tím, že jsou ty výpošty správné,
a bála se, že to udělá mnohem větší buch, než to nakonec udělalo.
A stejné obavy si děláme i teď.
Nakonec atmová energie přinesla, myslím, i mnoho dobrého,
ale pořádoufám, že to stejně nebo podobné bude i s umělu inteligenci.
No tak, či onak v Itálii už omezují po užívání GPT
asi nejviditelnějšího z těch jazykových modelů, současnosti z těch,
které mají přiné s nějakou revoluci od větců i od známých osobností
z ní také výzvyk pozastavení vývoje AI a spoň tedy na několik měsíců.
Ten svět tedy není úplně tak načený z toho všeho, co AI dovede?
Ono je to několik témat na jednou v Itálii, bylo GPT nebo čet GPT pozastaveno kvůli tomu,
že podle názoru tamního regulátora neodpovídá Evropským pravidlům
pro nakládání s osobními údají.
Což si myslím, že vlastně legitýmní postřech
lidi si často neuvidumí, že to, co napíšou do toho promptovatího okénka,
může někdo vědět, ono nejspíš nehrozí, že by se to propsalo do nějakých jiných odpovědí,
ale skrátka do databáze té služby se dostává spoustacitlivých údajů
a podle toho úřadu není jasné, jak s nimi provozovatelé nakládají,
takže to je asi oka připomínka, i když nevímysl je v pořádku samotné to řešení
skrátka na řízení vypnutí.
Italský úřad poukazuje mimo jiné na to, že před několika týdny unikli informace

o uživatelích toho toho četovacího robota.
Italia firma ČAT GPT,
linteligenc artificiale generativa,
ke generačuje imagini, videotesty, disenso kompiuto dal nulla.
A potom jsme viděli minulý týden otevřený dopis řady lidí,
kteří se umělou inteligenci akademické nebo je podnikatelský zabívají
a tě v něm navrhuji pozastavení vývoje všech modelů,
které jsou výkonější a mocnější než GPT-4 na 6 měsíců
a chtějí těch 6 měsíců využít k tomu,
aby se přijali nějaké všeobecně platné normy
a protokoli, proto aby ty služby a ty modeli byli bezpečnější
a neohrozili lidstvo takovým nebo makovým způsobem.
A za tou výzvou stojí poměrně známí lidé, ne?
Mask, vozniák, harary?
Přesně tak, všechni tě lidé nějakým způsobem
už se umělou inteligenci v minulosti zabívali konec konču.
Jestli se nepletu, tak i maskové peníze byly v začácích organizace OpenAI,
a i která je provozovatelem čet GPT.
No a není tedy teď tak trochu správná ta kritika
právě za tou výzvu směrem na maska,
že chce tu konkurenci zabrzdit, aby on sám mohl pracovat na něčem podobne?
Určitě to není, připomínka jenom, k maskový mezi těmi segnatáři
je i provozovatel stable diffusion, což je služba pro generování obrázku,
o které se říká, že teď v posledných týnech a měsících trošku stráci tempo.
Na druhou stranu bych se nerad zasekel na těch letech argumentech at hominem,
protože opravdu začínáme si zahrávat s něčím, co prekonává naše očekávání,
by to v tomhle kamžiků zřejmě není nějak zásedným způsobem nebezpečné,
a ten návrch na šestim měsíčních, řekněme pauzu nebo moratorium, je legítimní.
Na druhou stranu mě vlastně přijde ještě zajímavýši kritika
s opačné strany, v magazinu Time všel minulý týden článek Elizra Jadkowského,
který je intelektuálem zabývajícím se už dlouho umělou inteligencí,
a ten píše, že těch šest měsíců vlastně vůbec nic nevěří,
že obcházíme nějakou pandořinu skřinku, kterou když otevřeme,
tak už je pozdět okol je řešit a navrhuje nesmírně tvrdý postup
v podstatě ukončení jakéhokoliv dalšího vývoje s několika drobnými výmkami
a dostává se tam k takovým navrhum,
jako je bombardování datových center v případě,
že někdo tady tenhle ten zákaz překročí.
Já jsem asi v životě neviděl alarmistíčtější tekst,
někoho, kdo je zároveň v celku uznávaným odborníkem
a má tušení o tom a o čem píše.
Pokud z pravodejské služby tvrdí, že nějaká země,
která není součástí dohody, buduje počítačovou farmu,
kde se vyvíjejí ty nejmocnější umělé inteligence,

bojte se míň vyvolání mezinárodního konfliktu,
než porušení moratoria.
Buďte připravení vybombardovat nepřátelské datové centru.
Takže alarmistický, ale relevantní podle tebe?
Já věřím tomu, že to relevantní není,
ale ve chvíli, kde se bavíme o tom, že jednou z možností je,
řekněme, kolab s civilizace nebo nějakým jiným způsobem zálních listvat,
tak jaká pravdě porobnost nebo jaké rizikoje příliš malé na to,
abychom ho přijali.
Já sam alarmista nejsem, myslím si, že umělá inteligence toho
teď přináší víc dobrého než zlého, a že to ještě nějakou dobů,
tak bude, ale ve chvíli, kdy ty černé scenáře,
když jsou opravdu černé, tak je na místě poslouchat
a nepřicházit rychlé z oceutky.
Jak tom přidám ještě jeden názor, točiš vítěniš
z Mendelovy University v Brně tvrdí,
že ani sami vývojáři přesně neví, jak ty jazykové modely,
tedy on konkrétně tuším mluvil o GPT, jak dochází k závěrumke, kterým dochází.
Tak oni sami přiznávají, že ani oni úplně přesně nevědí,
jakým způsobem ten jejich produk, ten čet GPT, dojde k těm závěrumke, kterým dojde.
Tak nezní to trošku strašidelně?
Trošku strašidelně to může znít, zároveň je dobré připomenout,
že ať už se bavíme o jakékoliv kompleksní technologie nebo službě,
tak málo kdy dokáže jeden člověk popzat úplně celé její fungování.
Těchli těchle modelů dokážou popzat, jak se ty modely učili, na čem se učili,
znaj jejich možnosti jenom nedokážou u každé konkrétní odpovědě na každý konkrétní prompt
nebo tu otázku, kterou lidé té inteligenci zrávají vysvětlit,
proč to došlo přesněkní, protože je ta věc příliš velká a dojisté míry se učila sama.
Když vinelezneš něco jako čet GPT, co ti odpovídá na otázky a radí ti a pomáhá ti,
ale zároveň to může mít nějakou tu svou stěnou stránku,
ale ty vlastně nevíš, jak funguje ten mozek, tak to asi není úplně správně nebo?
Já si nemyslím, že by byla pravda, že nevíme, jak ten mozek funguje,
mi zkrátka jenom nejsme schopni vysvětlit každej jeho rozhodnutí, každý jeho krok,
ale pořád to ještě není nějaký mega kompleksní systém,
který by se sám upravoval tak jako třeba superpočítač golem tuším 14
v kni ze Stanislava, Lema, půlstletí staré. Tam ještě nejsme.
No jak to teda funguje? Jak ten mozek umělej inteligence funguje?
Pokud můžu nějakým způsobem rozumíme?
To rozumíme to. Ano, on funguje v zásadě mimořádně primitivně,
jenom na obrovských datech dopočítává, jaká další slova budu pravděpodobně následovat
potom promptu, potom zrání, které dostal jednoduchý příklad je,
když někám napíšeme játě dvě slova, tak za nimi může být miluju,
nenávidím na kopu, dlouho jsem ti neviděl.
Ta pravděpodobnost klesá uslov, jakuje televize, nebo ruchadlo

a ještě víc uslov v cizích jazyků.
I když ani tamto není úplně vyloučené, protože v těch učebných datech
mohl byť nějaký chybný článek, kde někdo uměle vymazal řádek
nebo nějaká absurdní poezie a konec koncu játě televize
by se asi taky někde mohl objevit, ale nejspíšne.
Vychvíli, kdy to zadání, kdy ten prompt rozšíříme o víc slov,
o nějaký kontext, kdy to bude, já tě mám místneši nomrát, játě,
tak tady je razatně stoupá pravděpodobnost, že zatím bude následovat slovo miluju
a klesá pravděpodobnost všech ostatních.
A teďka jenom, abych to jak schrnul prompt, zadám a tak dále,
ono to prostě funguje tak, že jaká si webová stránka,
do které ty můžeš psát nějaké úkoly pro tu, umělou inteligenci,
pro ten jazykový model a on na základě toho odpovídá ti do takového četu.
Třesně tak, v postatě kdo koli umí posílat, přijímat SMS-ky
nebo se s bavickámi rády, s příbuznými po nějakém messengeru nebo WhatsAppu,
umí ovládat, umělou inteligenci.
No a jak se teda ptát dobře, jak skládat ty otázky, ty prompty tak,
abych dostáhl těch, řekněme, nejlepších výsledků,
pokud si prostě chci zjednodušit život, ulechčit práci,
využít to dobře, co mi právě čet GPT a jiné Bing a tak dále na bízej.
Na mího pravdu důležité dodat, co nejvíc kontekstu.
Zároveň pamatovat na to, že jazikový model není nějaká znalostní databáze.
Když se koukám různě na sociálních sítích,
jak jsou lidi rozčarovaní z toho, co zadali a co jim to vyhodilo za hloupost,
tak jak to říztan, tam chybí nějaký praktický návod,
a v ní se často neuviromují, jaké jsou limity, které těch letěch modelů.
Je to část eště nedané tím, že jsme použijú to slovo napromptovaní
populádní kulturou.
50 letná svědesko-fantastické filmy připravovali na to,
že až přijde umělá inteligence, tak to bude něco,
jako počítař hal 9000 ve smírné odysy.
Nesmírně inteligentní, rafinovaný, můžeme říct, dvor,
nebo nějaká kvazy byto strá dokáže vyvraždit rafinovenými manipulacemi
posádku kosmitské lodi různě je manipulovat.
Ale ČGPT taková není. My jsme mýsli to nějaké
geniální profesorky dostely takového lehce
sětého lehkými drogami učitele ze střední školy,
který zkrátka řetězí slova, a často říká věci, které jsou nesmyslné.
Často seplete? Často seplete. Ano, požívá se pro to termín Halučinace,
protože mě napadla ta metafora těch lehkých drog.
Ve chvíli, kdy chceme, aby jsme dostávali relevantní odpovedi,
tak první základ je dodat co nejšerší kontext.
Když bys mě řekl, že potřeboješ nějaké otázky pro přiští díl,
tak já budu vidět, jo, tohle Matis dozlasu, dělá vynu hradskou 12,

chodí tam takový a makový lidé, má to nějaký formát,
ČGPT nic takového neví ani další modely,
potřeboje vysitli takovému dítěti, co přesně chceme,
jak to má vypadat, kdo jsme, pro kohotu děláme,
ideálně taky můžeme postavit nebo dát ČGPT nějakou roli.
Já jsem si tady vytyskil prompt, který využívám při editování
článku datového týmu českého rozlasu,
možná ukáže, jak to může vypadat.
Ještě předtím, když tam možná vložím ten článek
a dám nějaké zarání, tak dodán tenhle kontext.
Si zkušená editorka z pravodajství,
vystudovala si fakultu humanitních studií Univerzity Karlovy.
Přes Erasmus si strávila dva semestry v paříže i na Sorboně,
přes fulbrightou stipendium si pobívala v New Yorku.
Maj zase byl stáže v novinách Lemon a New York Times.
Ponávratu si byla 20 let redaktorkou,
šef redaktorkou a editorkou, si vzdělaná,
zkušená, sečtila, inteligentní, máš cít pro jazyk,
vždy volíš nejetyštější řešení, záležitě na tom,
aby byli teksty zrozumitelné a čtivé.
A teprv potom tam zarám ten článek
a zeptám se třeba na to, co v něm přebývá
nebo jakýby mohl mít titulek.
Počkej, takže Michale, ty nejenom vytvoříš cívíčko
protu umělou inteligenci, ale ještě ji lichotíš?
Jo, a lichotími záměrně, protože jsem,
a nejenom já, ale ukazuje se, že ve chvíli,
kdy my tu inteligenci nasměrujeme k tomu,
aby tu odpověď lovila v tekste,
kde k sobě lidi mluví slušně
a mají nějaké znaky inteligence,
tak ta odpověď je skrátka lepší.
Jo, protože ona potom sahá do hlouby internetu,
aby měla ta zdrová data a nějakým způsobem ti odpovídá,
takže ty chceš už tím promptem, tím zadáním nasměrovat,
aby sahala do správné škatulky.
Aby to skrátka neleslo z nějakých hádek podčlánky,
kde se lidi vulgárně urážijí a jedou nějaké hity.
Tam ještě důležité zmínit,
že ta treningovan, bo ta učibní data, na kterých GPT,
če GPT funguje, konči někde v roce 2021,
tož částešně opět přispívá k těm častím halucinacím.
Když se jí zeptáte, kdo vládne, velké Britány,
nebo na průběch invaze Ruské federace na Ukrajinu,

tak ona o tom vůbec neví, tam informace nejsou,
může si něco domyslet, ale často si domýšlí věci,
které jsou až legrašně chybné.
Oná si potom zátěmy chybami stojí, jako že ví, že chybuje?
Záleží. Záleží.
Zrovna, když k té inteligenci přistupujeme přestočitové rozraní,
tak tam je nějaký systémový prompt,
který velí být taková pokorna,
a když jí upozorníme na to, že něco špatně,
tak si to pokůší opravit.
A ohledně toho, o čem stady bavíme.
Tak k čemu vše mu se to dá teda používat?
Kdo teď ty prompty, které si tady nadhazoval,
které by mohli fungovat v tom jazykové modolu,
kdo je může všechno využívat?
Reálně třeba, jsem programátor.
Už mě nebavípšá takové nějaké základní kody.
Napíše to, čet GPT, Bing a další za mě?
Napíše. Právě proto, že se učilo z velké části na internetových stránkách,
na těch zpousta ukázek kodu,
včetně popisu, co ten kod má dělat, nebo co naopak nedělá v čem je problém.
Sámi takhle využívám u takových těch repetitivních činností,
které nejsou moc na přemýšlení,
ale spíš na to, že si musím pamatovat nějakou syntaxi,
kterou si nepamatuju typicky otevřituhle složku
a načti všechny tabulky a spoje do jedné tabulky.
Tak je to otázka jednořádkového promptu
a do několika sekund dostanu kod, který výječinou funguje.
Že to je genialní encyklopédy o tom není sporu?
Co překlady třeba?
No pardon, já bych si o tom na trošku...
Tak hádej se dovidí.
Protože my opravdu nemůžeme spolehat na to,
že ty informace jsou věcně správné, které dostaneme na spátek.
Připomínám, že je to jazykový model,
který skrátka doplňuje nejpravděpodobnější pokračování,
což ne, vždycky je správné pokračování
a v tuhle chvíli v tom systému není mechanizmus,
který by tu správnost dokázal zajistit.
Praxi jsem došel k tomu, že ty odpovědí bývají super
a dá se na ně docela spolehat,
když chceme nějaké spíš obecné věci
popsať rozdíl mezi naftovým a benzínovým motorem.
K tomu existuje spoustatekstu, je to v celku jednoduchá věc.

Můžeme se i hezky volit, jak složitě to chceme mít vysvětlené,
můžeme tam dopsat, že to chceme mít jako pro pětile té dítě,
nebo jako pro studentku vysoké školy.
V tom je dobrá.
A ve chvíli, kdy jdeme po nějakých tětěrných detailch,
typu, co měli na hlavě účasnici nějaké bitvy ve třetím století,
jestli to byli čepice, přilby, nebo tak tak nás.
Může návé s dobrým sněrem, ale taky může v velice suveréně lhát.
A pak je to dále hrozně složité toho do halitne,
nebo můžu už tím promptem nějakým způsobem záručit,
že mlhát nebude asi pouze jenom tím kontextem,
že bych ji řekl, sež úspěšný historik.
Přesně tak, já si narazel na jednu šablonu promptu,
tuším, že na blogu LessWrong.
A ten prompt vlastně staví takovou situaci z nějaké beletrie,
ve které je postava nějaké ženy,
která vůbec nemá smysl pro humor, nikomu by nikdy nelhala,
je strašně chytrá a snaží se všem odpovídat správně,
a kamaráci ji na něco zeptá.
To je ten náš dotáz, co chceme vědět.
A její odpovět je, dvojitečka.
A zatím můžeme doufat, že bude s větší pravděpododností
správná odpovět, než bychom tady ten úvod neměli.
No a když se teda vrátím k té své původní otáce,
to těž co třeba překladatelství, překlady.
Můžu se bavit, co mělo inteligencí v Češtině, v Angličtině
a dokážeme si těmi jazyky i překládat?
Dokáže překládat, to je důležité říct,
my se můžeme ptát rovnou normálně Česky.
U většiny témat mi přijde, že o něco lepší odpovědi dává v Angličtině
myslím ve smyslu propracovanější, srozumitelnější
a to prostě proto, že toho učibního materiálu
v Angličtině tam bylo nesrovnatelně víc.
No a překlady, vlastně není moc nová věc,
my třeba máme už několik let nástroj D-Pool,
který podobným způsobem překládáme z různými jazyky
a překládá docela slušně.
Kde to posouvá ČGPT dál je,
že s ním můžeme přímo konverzovat, což mě přijde skvělý,
ve chvíli, kdy se třeba učíme nějaký nový jazyk,
tak pořáda si dává smysl chodit na lekce se rodili mluvčím,
ale jenom si tak nasimulovat, že se bavíme s někim nějaké kavárně,
co si objednáme a kolik to bude stát, tak mě to přijde hrozně super.
Může to být v Angličtině a v desítkách dalších jazyků

včetně několika umělých.
A pak jsou tu ty generátory obrázku.
Je to stable diffusion, je to mid-journy, je to další a tak dále.
Já si říkám, jestli teda přesty jazykové modely,
si nemůžu tak napromptovat to, co bych pak chtěla,
by ten generátor obrázku mi vytvořili,
si mi nemůže ta uměla inteligence pomoct i v tomhletom.
A býhledově to určitě možné bude,
ale připomínám, že cvičební data ČPT končí v roce 2021
a tehdy ještě tyhle nástroje nebyly veřejné,
ty je na generování obrázku.
A tím pádem na internetu a v dalších materiálech
nikde nebyly pobsané dobré prompty,
ale pokud si někdo chce generovat obrázky,
ilustrace, fotografie v těchle nástrojích,
tak se dá doguglovat,
ale se se vracíme k tomu tradičnímu internetu spoustá materiálu,
kterým se říká promptbuky a podobnými vyrázy,
kde se lidi dělí o dobré praxe.
A ta dobrá praxe konec konců vypadá dost podobně jako učet GPT.
Pamatujeme na to, že ten nástroj neví pro kogo to, co chceme generuje
a funguji takové věci, které by lidskou grafičku mohli až urazit.
Když doplníme, že ten obrázek má být hezký,
tak bude spíš hezký, než když to tam není,
protože se to opět učilo na milionech nějakých popsaných obrázku
a jsou tam i ošklive.
Potřebuje to trošku pošťouchnout tím směrem, kterým chceme.
Mým chodem neděsí tě, jak reálně ty obrázky vypadají,
jeden za všechny popeš v bílé pérovce.
Teď jsem četl ve Washington Postu,
že to, co nezvládali ty generátory, to znamená trá prsty úrokou,
že to už se taky jako dá.
Zlepšuje se to ten pokroke vidět každý měsíc.
Tal se si sly mě to neděsí, tohle mě vlastně ani tak neděsí,
protože my už dlouho žijeme s technologiami,
které nám umožňují dělat poměreně fotorealistické obrázky,
věcí, které se nestali.
Ve 40. a 50. letech si stali nechával odmazat
fotografii popravené ministry, potom přišel v 90. letech Photoshop.
Teď je na tom asi trochu zneklidňující,
že tyhle nástoje má v rukou každý,
ale to taky znamená, že možná budeme všichni všeobecně ostražitější
k tomu, co vidíme na internetu.
Přiznám si, že sám jsem víc nepokojený čevatím,

jak je stále snaší na podobitni čí hlas,
zase podobnými nástoji strojového učení umělu inteligencí,
ve chvíli, kdy mě bude v noci s cizího čísla
volat někdo, kdo zní úplně stejně jako můj blízký člověk
a bude mít i podobné vyjadřování, protože všechno tohle jsou věci,
které je do poměrně snadnost internetu nebo nějakou lustí zistit
a bude poměchtit poslat hrychlé penízení kám, protože ho unesli,
tak to asi neudělám, doufám, že to neudělám,
ale zkrátka těch možností manipulace přichází mnohem víc,
než jsou jenom falečné obrázky.
Nemyslíš si, že si můžeš dobudouc nastát odborníkem na prompty,
nějakým si prompt-engineerem,
není možné, že nějaká taková nová profesy vznikne?
Tak bych zás připoměl Google a to, že nemáme profesy Google'arů profesionálních,
všichni nějak vyhledáváme na internetu ve své práci i ve volném čase,
myslím, že se něco porudného stane z prompty
a myslím, že se to stane ještě letos,
že většina lidí si zkrátka vytvoří nějakých pár svých promptů,
které jim pomáhají řešit to, co je v práci nebo po práci,
nejvíc držuje, otravuje, nebaví, chtějí to mít rychlé skrku
a zřejmě velké instituce budou mít někoho, kdo jim v tom pomáha,
trošku to zdokonaluje nebo v dedatabázi nějakých osvečených šablon,
ale nemyslím si, že by to samo o sobě bylo nějakým masovým povoláním.
A bude ještě letos pokračovat vývoj jazykových modelů
a umělé inteligence, nebo dojde k té stopce, po které někteří volají?
Nevím a nemůžu vědět, ale co mě přijde v tuhle chvíli ještě zajímavější,
než je vývoj těch velkých modelů, jako je GPT,
které, a to tu myslím ještě nezaznělo běží v nějakém klaudu
na obrovské vypočetní kapacite, tak mě přijde hrozně zajímavé,
že přicházejí první modely, které se dají provozovat na běžných,
běžných, byť draších, noudbůcích a pracovních počítačích.
To znamená, že i kdyby se lidstvo schodilo na tom,
že tahle technologie vede ke zkáze a že s ní musíme zatočit,
tak v tom vodu, ve kterém jsme teď už nejspíš zůstaneme na vždycky,
protože by ta stopka vyžadovala konfiskaci počítaču,
zjednodušeně řečena.
Tak moc díky, že jsme o tom mohli mluvit.
Já děkuji za pozování, krásný den.
Tohle už je všechno z Vinohradské 12 z pravodejského podkástu Českého rozhlasu.
Dnes s Michalem Kašpárkem mým kolegou datovým novidářem Českého rozhlasu,
který se zajímá o klady i zápory umělej inteligence
a nejrůznějších takzvaných jazykových modelů nebo neuronových sítí.
Tahle epizoda Vinohradské 12 je ukonce a to znamená jediné,
totiž už připravujeme další najdeteji jako vždy na irozhlaz.cz

v aplikaci Můjrozhlaz a ve všech dalších podkástových aplikacích.
Naslišenou zítra.