

Reasoning AI modely, jejich prompting a Deep Research nástroje

1. Úvod do reasoning modelů

Reasoning modely jsou velké jazykové modely (LLM) navrženy tak, aby před odpovědí „přemýšlely“ ve více krocích. Na rozdíl od klasických LLM (jako GPT-4 nebo ChatGPT) se nesnaží ihned predikovat finální odpověď, ale rozkládají komplexní problém na dílčí **myšlenkové kroky** (tzv. *chain-of-thought*) ([A Visual Guide to Reasoning LLMs - by Maarten Grootendorst](#)). Tím se více blíží lidskému uvažování a dokážou řešit obtížnější úlohy, zejména v matematice, logice či programování. OpenAI tuto novou řadu modelů (označovaných o1) trénuje pomocí posilovaného učení tak, aby si model „tříbil“ své myšlenkové postupy, zkoušel různé strategie a uměl rozpoznat a opravit vlastní chyby ([Introducing OpenAI o1 | OpenAI](#)). Výsledkem je, že reasoning model dokáže vyřešit mnohem složitější problém než klasický model – například v testu z mezinárodní matematické olympiády správně vyřešil 83 % příkladů, zatímco původní GPT-4o jen 13 % ([Introducing OpenAI o1 | OpenAI](#)).

([A Visual Guide to Reasoning LLMs - by Maarten Grootendorst](#)) Schéma porovnávající klasický „regular“ LLM (vlevo) s „reasoning“ LLM (vpravo). Reasoning modely vkládají před finální odpověď sekvenci vnitřních **myšlenkových kroků** (červeně), tedy v podstatě „uvažují“ a kontrolují se, než odpoví ([A Visual Guide to Reasoning LLMs - by Maarten Grootendorst](#)).

Proces generování odpovědi: Klasický LLM jako ChatGPT generuje odpověď převážně jednofázově – vstupní otázku zpracuje jedním průchodem a ihned generuje finální text. Pokud je úloha složitá (např. logická hádanka), klasický model často potřebuje uživateli naznačit, aby uvažoval krok za krokem (např. *prompt* „Let’s think step by step“), jinak může „skočit“ k nesprávnému závěru. Naproti tomu reasoning LLM má **zabudované** vícefázové uvažování přímo v modelu ([Prompt Engineering for OpenAI’s O1 and O3-mini Reasoning Models | Microsoft Community Hub](#)). To znamená, že model si interně vytváří a prochází řetězec dílčích inferencí (aniž by je nutně zobrazil uživateli) a teprve poté formuluje závěr ([A Visual Guide to Reasoning LLMs - by Maarten Grootendorst](#)). Díky tomu umí důsledně analyzovat zadání, zvolit strategii řešení a sám sebe kontrolovat v průběhu generování. Modely OpenAI o1 byly trénovány právě k tomu, aby trávily více času „přemýšlením“ před odpovědí podobně jako člověk – během nácviku se učí zpřesňovat postup řešení a rozpoznat, kdy se ubírají chybným směrem ([Introducing OpenAI o1 | OpenAI](#)). Tento proces je sice pomalejší, ale výrazně snižuje výskyt velkých chyb u obtížných problémů (o1-preview měl o 34 % méně závažných chyb než původní GPT-4) ([OpenAI o1 explained: Everything you need to know](#)).

Vhodnost reasoning modelů pro různé úlohy: Díky schopnosti vnitřního uvažování excelují reasoning LLM zejména v **komplexních úlohách**, které vyžadují logické odvození, vícestupňový postup nebo detailní analýzu. Patří sem například:

- **Matematické a logické problémy:** Např. soutěžní příklady z fyziky, chemie či matematiky, komplexní slovní úlohy. (GPT-4o jich vyřešil jen ~13 %, zatímco reasoning model 83 % ([Introducing OpenAI o1 | OpenAI](#)).)
- **Programování a debugging:** Řešení složitých programátorských úloh a ladění kódu. Reasoning model dosahuje vysokých skóre v coding soutěžích (např. 89. percentil na Codeforces) ([Introducing OpenAI o1 | OpenAI](#)) a dokáže krokově rozebrat kód i najít chyby.
- **Vědecké a technické analýzy:** Anotace genomických dat, odvozování vzorců v kvantové optice apod., kde je potřeba logicky zkombinovat více faktů ([Introducing OpenAI o1 | OpenAI](#)).
- **Rozsáhlé úlohy s dlouhým kontextem:** Reasoning modely podporují extrémně dlouhý vstup (OpenAI o1 až 128k tokenů) ([Prompt Engineering for OpenAI’s O1 and O3-mini Reasoning Models | Microsoft Community Hub](#)), takže zvládnou vstřebat a analyzovat např. celou kapitolu textu nebo výsledky vícero experimentů najednou.

Naopak pro **běžné a přímočaré úkoly** nemusí být reasoning model vždy ideální volba. OpenAI uvádí, že v mnoha běžných případech (např. konverzace, všeobecné znalostní dotazy nebo kreativní psaní) je standardní GPT-4o zatím výkonnější a praktičtější ([Introducing OpenAI o1 | OpenAI](#)). Klasické modely excelují v:

- **Široká znalost tréninkových dat:** GPT-4 má obrovskou obecnou znalost světa a umí plynule generovat text. Umí dobře reagovat na obecné dotazy, což reasoning model (specializovaný na řešení) nemusí – např. o1-preview nedokázal odpovědět na otázku o sobě samém kvůli užším znalostem mimo tréninkové domény ([Prompt Engineering for OpenAI’s O1 and O3-mini Reasoning Models | Microsoft Community Hub](#)).
- **Konverzace a kreativní obsah:** ChatGPT (GPT-4) je vyladěn pro přátelskou konverzaci, vysvětlování v různých stylech či tvorbu příběhů. Reasoning modely se soustředí na analytickou správnost, tón jejich odpovědí bývá spíše suchý a technický. Například pro napsání eseje nebo bajky je klasický model vhodnější.
- **Rychlé odpovědi:** U jednoduchých otázek (např. faktických) dá ChatGPT odpověď během pár sekund. Naproti tomu reasoning model může zbytečně „přemýšlet“ a zdržovat. U OpenAI o1 byly zavedeny ukazatele průběhu, protože některé dotazy mu trvají znatelně déle ([OpenAI o1 explained: Everything you need to know](#)). Pro úlohy, kde je rychlost důležitější než dokonalá hloubka, je lepší klasický model.

Níže je shrnutí vhodnosti/nedostatků pro různé typy úloh:

Typ úlohy	Klasický LLM (GPT-4)	Reasoning LLM (o1 a pod.)
Komplexní matematické úlohy	Omezený výkon – GPT-4 vyřešil jen 13 % příkladů z IMO kvalifikace ([Introducing OpenAI o1	OpenAI](https://openai.com/index/introducing-openai-o1-preview/#:~:text=In%20our%20tests%2C%20the%20next,in%20our%20technical%20research%20post)).
Soutěžní programování	Velmi dobrý (GPT-4 dosahuje ~20 % na Codeforces)	Vynikající – o1 dosahuje 80–90 % na programátorských benchmarcích (OpenAI o1 explained: Everything you need to know).
Obecné znalostní dotazy	Excelentní – široká znalost díky tréninku na webových datech.	Může postrádat kontext mimo svou doménu (nutno dodat info v promptu) ([Prompt Engineering for OpenAI's O1 and O3-mini Reasoning Models
Konverzace a kreativní psaní	Přirozené a kreativní výstupy, vyladěné pro dialog.	Formální a analytický styl, zaměřený spíše na přesnost než kreativitu.
Rychlost odpovědi	Rychlá – obvykle v řádu sekund.	Pomalejší – „uvažuje“ déle, u složitých dotazů minuty (má progress bar) (OpenAI o1 explained: Everything you need to know).
Aktuální znalosti	Bez nástrojů nezná data po 2021; s web browsing pluginem může vyhledat info.	Zatím nemá přímý přístup k webu (o1-preview neuměl prohlížet web ([Introducing OpenAI o1

Historie a přehled reasoning modelů: První LLM (GPT-3, GPT-4) ukázaly překvapivé emergentní schopnosti řešit složité úlohy, ale často **halucinovaly** a „přeskočily“ důležité kroky. Výzkumníci proto přišli s technikami jako *Chain-of-Thought prompting* (2022) – kdy se modelu nařídí „rozmysli si to krok za krokem“, což zlepšilo úspěšnost složitých úloh ([Prompt Engineering for OpenAI's O1 and O3-mini Reasoning Models | Microsoft Community Hub](#)). V roce 2023 se objevily další postupy (self-consistency, tree-of-thought, atd.), ale šlo stále o *triky* v *promptu* pro klasické modely.

Zlom nastal v roce 2024, kdy se objevily modely explicitně trénované k uvažování:

- OpenAI o1 (2024)** – první veřejně dostupná série reasoning modelů. *O1-preview* vyšel v září 2024 ([Introducing OpenAI o1 | OpenAI](#)), navázal o1 v prosinci 2024 jako součást ChatGPT Pro. Tyto modely mají zabudovaný chain-of-thought a dosahují špičkových výsledků v STEM oblastech ([Introducing OpenAI o1 | OpenAI](#)) ([OpenAI o1 explained: Everything you need to know](#)). O1 umí i analyzovat obrázky a má kontext 128k tokenů. K dispozici je i **o1-mini**, menší a levnější varianta zaměřená na rychlé programování (80 % levnější provoz než o1) ([Introducing OpenAI o1 | OpenAI](#)). Nevýhodou o1 je zatím omezená dostupnost (jen pro platící uživatele/API Tier 5) a vyšší cena; navíc v rané verzi neměl integraci nástrojů jako prohlížení webu či nahrávání souborů ([Introducing OpenAI o1 | OpenAI](#)).
- Google Gemini (2024)** – nová generace modelů od Google DeepMind. Gemini 1.5 Pro, spuštěný v prosinci 2024, obsahuje *agentní režim Deep Research* (viz kap. 3) a kontext až 1 milion tokenů ([Gemini: Try Deep Research and Gemini 2.0 Flash Experimental](#)). Gemini

klade důraz na *vícezkrokové plánování* – při komplexním dotazu nejprve navrhne plán řešení, který může uživatel upravit, a poté iterativně prohledává web a skládá poznatky ([Gemini: Try Deep Research and Gemini 2.0 Flash Experimental](#)). První verze měla nižší faktografickou přesnost (cca 6 % v benchmarku Humanity's Last Exam ([Perplexity Deep Research Takes on OpenAI & Gemini](#))), ale díky silné integraci s Google vyhledáváním a agentní architekturou má velký potenciál. Dostupný je v rámci služby **Gemini Advanced** (~20 USD měsíčně) v angličtině ([Perplexity Deep Research Takes on OpenAI & Gemini](#)).

- **Open-source projekty (2024)** – komunitní úsilí vedlo ke vzniku open-source reasoning modelů. Významným počinem je **DeepSeek-R1 (2024)** – otevřený model trénovaný pomocí RL, který se výkonem blíží proprietárním modelům ([A Visual Guide to Reasoning LLMs - by Maarten Grootendorst](#)). DeepSeek využívá architekturu Mixture-of-Experts (MoE) a dosáhl značných úspěchů např. v programátorských úlohách (51,6 % na Codeforces vs. GPT-4 jen 23,6 % ([DeepSeek](#))) a v matematickém testu AIME (39 % vs. GPT-4 9 %) ([DeepSeek](#)). Tvůrci jej uvolnili zdarma (váhy modelu i API), takže jej lze nasadit bezplatně ([A Visual Guide to Reasoning LLMs - by Maarten Grootendorst](#)). Omezením může být složitější použití (je nutné vlastní HW pro 70+ miliard parametrů) a mírně nižší spolehlivost či znalost světa ve srovnání s obřími uzavřenými modely.
- **Další modely:** K dalším patří např. **Anthropic Claude 2 (2023)** – ten sice nebyl speciálně trénován RL na reasoning, ale díky 100k kontextu a vylepšenému tréninku také umí řešit složité úkoly. Elon Musk představil **xAI Grok (2023)** s přístupem k internetu, avšak v testech zatím značně zaostává (Grok-2 dosáhl jen ~3,8 % v HLE benchmarku ([New OpenAI 'Deep Research' Agent Turns ChatGPT into a Research Analyst -- Campus Technology](#))). Dá se očekávat, že i další hráči (IBM, Meta aj.) zahrnou prvky reasoning do svých budoucích LLM.

Rekapitulace: *Reasoning AI modely* představují nový přístup, kdy modely při inferenci simulují analytické myšlení v krocích. Díky tomu dokáží s vysokou přesností řešit komplexní úlohy v matematice, programování či vědě, kde klasické LLM narážely na své limity. Za cenu vyšších výpočetních nároků a pomalejší odezvy přinášejí dramatické zlepšení – např. v testech překonávají GPT-4 o desítky procent ([Introducing OpenAI o1 | OpenAI](#)). Na běžné každodenní dotazy nebo kreativitu ale zůstávají praktičtější klasické modely. Vývoj reasoning modelů akceleroje (OpenAI o1, Google Gemini, DeepSeek aj.) a postupně se rozšiřuje jejich dostupnost. Pro vývojáře to otevírá nové možnosti, jak nechat AI promyslet složité problémy do hloubky a získat spolehlivější výsledky tam, kde dosud LLM selhávaly.

2. Prompting reasoning modelů

Abychom z reasoning modelů dostali maximum, je třeba přizpůsobit styl vytváření dotazů neboli **prompt engineering**. Nejprve si připomeňme základní pravidla promptování klasických LLM a poté zdůrazníme odlišnosti u reasoning modelů.

Základní pravidla prompt engineeringu (pro klasické modely): U běžných LLM, jako je ChatGPT, platí několik osvědčených zásad ([AI Prompt Best Practices: Learn Prompt Engineering](#)) ([AI Prompt Best Practices: Learn Prompt Engineering](#)):

- **Jasnost a konkrétnost:** Formulujte dotaz jednoznačně. Vyhněte se vágním formulacím. Např. místo „Řekni mi něco o AI“ raději „Vysvětli hlavní rozdíly mezi umělou inteligencí a strojovým učením v ~300 slovech“. Čím specifičtější popíšete požadovaný obsah (téma, kontext, perspektiva) a formu (např. „v bodech“, „souhrn o 200 slovech“), tím lépe model porozumí zadání ([AI Prompt Best Practices: Learn Prompt Engineering](#)).
- **Poskytnutí kontextu:** Model nemá trvalou paměť konverzace ani znalost vašeho záměru, proto mu dodejte veškeré podklady, které potřebuje. Je-li dotaz navázán na předchozí diskuzi, připomeňte klíčová fakta. Např. „Navážeme na předchozí text: [shrnutí]. Na základě něj odpověz...“ Tím modelu jasně vymezíte východisko ([AI Prompt Best Practices: Learn Prompt Engineering](#)).
- **Strukturování a formát:** Řekněte modelu, v jakém formátu má odpovědět. Můžete výslovně uvést „odpověz formou očíslovaného seznamu“ nebo „vrať JSON s následujícími poli“. Také můžete strukturovat samotný prompt – např. odrážkami vypsát požadavky. Dobrým trikem je nastavit roli: „Představ si, že jsi zkušený lékař...“, což modelu napoví tón a styl odpovědi ([AI Prompt Best Practices: Learn Prompt Engineering](#)) ([Prompt Engineering for OpenAI's O1 and O3-mini Reasoning Models | Microsoft Community Hub](#)).
- **Iterace a vylepšování:** Skvěle zformulovat prompt na první pokus se podaří málokdy. Berte to jako iterativní proces: *zadání* → *výstup* → *úprava zadání* → *lepší výstup*. Pokud odpověď modelu není podle představ, upřesněte dotaz, přidejte omezení nebo rozdělte problém na menší části a zkuste to znovu ([AI Prompt Best Practices: Learn Prompt Engineering](#)). Model je citlivý na formulace, někdy pomůže jen drobná změna formulace či otázky.
- **Odstranění dvojznačnosti:** Identifikujte potenciálně nejasné pojmy a upřesněte je. Např. dotaz „Řekni něco o bankovních vkladech“ je nejednoznačný (banka vs. říční břeh v angličtině *bank*). Lepší je „Vysvětli, jak fungují bankovní vkladové účty v osobních financích“. Pokud je potřeba určit perspektivu (právní, historickou) nebo co vynechat, uveďte to. Také se vyplatí zmínit, co model *nemá* zmiňovat, např. „Neuváděj technický žargon“ ([AI Prompt Best Practices: Learn Prompt Engineering](#)).
- **Příklady (few-shot):** Pokud je úkol složitý nebo formát neobvyklý, můžete modelu ukázat jeden dva příklady vstupu a požadovaného výstupu (*few-shot learning*). Např. „Příklad: Vstup X -> Odpověď Y.“ To dá modelu vzor, kterého se držet. U klasických modelů to často pomůže zvýšit přesnost interpretace zadání. (U reasoning modelů naopak viz níže.)

Tyto zásady platí obecně pro všechny LLM. **Reasoning modely** však mají některé zvláštnosti, které je dobré při promptování zohlednit ([Prompt Engineering for OpenAI's O1 and O3-mini Reasoning Models | Microsoft Community Hub](#)) ([Prompt Engineering for OpenAI's O1 and](#)

O3-mini Reasoning Models | Microsoft Community Hub):

- **Vestavěné řetězení myšlenek:** U reasoning modelu **není nutné ho pobízet**, aby uvažoval krokově – dělá to sám automaticky. Například není potřeba přidávat fráze typu „*Pořádně si to promysli*“ nebo „*krok za krokem*“. U modelů OpenAI o1/o3 by to bylo plýtvání tokeny a může to být i kontraproduktivní ([Prompt Engineering for OpenAI's O1 and O3-mini Reasoning Models | Microsoft Community Hub](#)). (Tyto pokyny mají smysl u GPT-4o, který jinak multi-step režim nespustí ([Prompt Engineering for OpenAI's O1 and O3-mini Reasoning Models | Microsoft Community Hub](#)).) **Stručný přímý dotaz často stačí**, model si interní logiku rozvede sám.
- **Nutnost dodat znalosti:** Reasoning modely (např. o1) mají užší záběr znalostí mimo svůj tréninkový fokus ([Prompt Engineering for OpenAI's O1 and O3-mini Reasoning Models | Microsoft Community Hub](#)). Nejsou připojené k internetu (pokud nepoužijete speciální nástroj) a mohou postrádat některá fakta, která by běžný GPT-4 znal. Proto **pokud se dotaz týká něčeho obskurního nebo velmi nového, poskytněte informace v promptu**. Např. chcete-li právní analýzu a model nemá v paměti daný zákon, uveďte text zákona jako součást zadání. U klasického GPT-4 byste to nemuseli – ten má širší encyklopedické povědomí. U reasoning modelu však raději nic důležitého nevynechávejte ([Prompt Engineering for OpenAI's O1 and O3-mini Reasoning Models | Microsoft Community Hub](#)) ([Prompt Engineering for OpenAI's O1 and O3-mini Reasoning Models | Microsoft Community Hub](#)).
- **Velký kontext využijte, ale nezneužíjte:** Tyto modely podporují extrémně dlouhé vstupy (stovky stran textu). To je skvělé pro komplexní zadání – můžete vložit všechny relevantní dokumenty, data atd. Dbejte však na relevanci: příliš balastu může model zmást nebo zbytečně zatížit ([Prompt Engineering for OpenAI's O1 and O3-mini Reasoning Models | Microsoft Community Hub](#)). Dejte jen informace, které opravdu potřebuje, a vynechte nerelevantní pasáže.
- **Minimum příkladů:** O1 a spol. obecně nepotřebují *few-shot* ukázky – jsou natrénované tak, aby úlohu pochopily i z popisu. Doporučení je začít tzv. *zero-shot*, tedy jen jasně popsat úkol bez příkladů ([Prompt Engineering for OpenAI's O1 and O3-mini Reasoning Models | Microsoft Community Hub](#)). Pokud by model zcela nepochopil formát, stačí dát jeden jednoduchý příklad. Rozhodně ale **nedávejte dlouhé řetězce příkladů**, může to naopak *zhoršit výkon* (model začne spíš napodobovat příklad než samostatně řešit) ([Prompt Engineering for OpenAI's O1 and O3-mini Reasoning Models | Microsoft Community Hub](#)). Například u právní analýzy není třeba v promptu demonstrovat kompletní vzorový rozbor případu – raději jen napište „*Použij formát IRAC (Issue, Rule, Analysis, Conclusion)*“ a model už odpověď podle toho uspořádá ([Prompt Engineering for OpenAI's O1 and O3-mini Reasoning Models | Microsoft Community Hub](#)).
- **Nastavení role a tónu:** I u reasoning modelu pomáhá úvodní systémová zpráva nastavit správný styl. Například „*Jsi zkušený právní analytik, který srozumitelně a logicky vysvětluje aplikaci práva...*“ pomůže modelu zvolit odpovídající formální tón ([Prompt Engineering for OpenAI's O1 and O3-mini Reasoning Models | Microsoft Community Hub](#)). Samotné logické uvažování by model zvládl i bez toho, ale role zajistí konzistenci výstupu (např. že bude citovat fakta, používat terminologii oboru atd.).
- **Specifikace formátu a rozsahu:** Tyto modely velmi dobře dodržují instrukce k formátování, takže toho využijte. Např. „*Seřaď výstup do tabulky*“, „*odpověz JSONem se strukturou...*“, nebo „*uveď odpověď v pěti odrážkách*“. Předejete tím nejasnostem ve výstupu ([Prompt Engineering for OpenAI's O1 and O3-mini Reasoning Models | Microsoft Community Hub](#)). Rovněž dejte najevo požadovanou podrobnost: „*odpověz krátce jedním odstavcem*“ vs. „*poskytni detailní rozbor v několika odstavcích*“ ([Prompt Engineering for OpenAI's O1 and O3-mini Reasoning Models | Microsoft Community Hub](#)). Model sám neví, kolik detailů chcete, ale pokyn dodrží.
- **Kontrola a ověření:** I reasoning model se může mýlit. U kritických úloh je dobré použít navazující dotaz pro ověření. Například se zeptejte „*Jsi si svým závěrem jistý? Vysvětli proč.*“ nebo nechte model rekapitulovat, jak k závěru došel. O1 má tendenci k **self-checking** – může sám upozornit, pokud mu něco nesedí ([New OpenAI 'Deep Research' Agent Turns ChatGPT into a Research Analyst - Campus Technology](#)). Toho lze využít: „*Zkontroluj, zda jsi použil všechna fakta a zda závěr opravdu vyplývá z práva.*“ Pokud něco opomněl, v následné odpovědi to přizná a doplní ([Prompt Engineering for OpenAI's O1 and O3-mini Reasoning Models | Microsoft Community Hub](#)). Pro naprostou jistotu lze spustit model vícekrát a porovnat výsledky (konsistence napovídá spolehlivost). Samozřejmě finální ověření citlivých výsledků by měl udělat člověk.

Shrňme **nejlepší praktiky promptingu reasoning modelů** do přehledu ([Prompt Engineering for OpenAI's O1 and O3-mini Reasoning Models | Microsoft Community Hub](#)) ([Prompt Engineering for OpenAI's O1 and O3-mini Reasoning Models | Microsoft Community Hub](#)):

- **Formulujte úkol jasně a přímo**, zbytečně ho neobalujte okolo. Reasoning model nepotřebuje tak bohatý kontext konverzace jako ChatGPT – stačí mu konkrétní otázka/problém.
- **Dodejte všechny klíčové informace**, které model potřebuje k vyřešení úlohy (zvlášť pokud jsou specializované či aktuální). Naopak vynechte irelevantní text, který by ho rušil.
- **Nepoužívejte dlouhé prompt-cesty a příklady**, pokud to není nutné. O1 se nespolehá na kopírování příkladů, umí si poradit *zero-shot*. Jeden příklad maximálně pro upřesnění formátu.
- **Specifikujte požadovaný formát a detail**, model to dodrží. Nastavte roli nebo styl, má-li být odpověď např. formální zpráva, seznam kroků, kód funkce apod.
- **Nevnucujte mu chain-of-thought v promptu** – u reasoning modelu to není třeba a můžete tím přijít o cenné tokeny ([Prompt Engineering for OpenAI's O1 and O3-mini Reasoning Models | Microsoft Community Hub](#)). Naopak u klasického GPT-4 byste to dělali (aby vůbec uvažoval).

- **Iterujte a testujte:** I když o1 zpravidla zvládne komplexní úlohy napoprvé, nebojte se prompt doladit, pokud výsledek není podle představ (např. přidat omezení délky). Malá změna formulace může výrazně zlepšit výsledek.
- **Ověřujte důležité výstupy:** Při nasazení v praxi zajistěte, aby kritické odpovědi prošly verifikací. Využijte model pro sebekontrolu (follow-up dotaz), a samozřejmě finální expertní kontrolu.

Ukázkové prompty a jejich vysvětlení:

Příklad 1 – Právní analýza:

Uvažujme, že potřebujeme od modelu posoudit, zda určitá situace naplňuje znaky porušení smlouvy. **Špatný prompt (pro GPT-4o):** „Je A odpovědný za porušení smlouvy?“ – je příliš stručný a klasický model by mohl halucinovat. **Lepší prompt (pro reasoning model o1):**

System: Jsi právní analytik, který rozhoduje spory dle amerického smluvního práva.

User: Strany uzavřely smlouvu o dodávce zboží. A se zavázal dodat 100 ks do 1. května, za což mu B zaplatí 10 000 USD. A dodal jen 80 ks a zbytek nedodal vůbec. B zaplatil jen 8 000 USD.

Otázka: Porušil strana A smlouvu a může po ní strana B žádat náhradu škody? Svě závěry vysvětli logicky a použij formát IRAC (Issue, Rule, Analysis, Conclusion).

Tento prompt jasně uvádí fakta (co kdo udělal) v odděleném odstavci, čímž modelu poskytuje výchozí **kontext**. Následuje explicitní **právní otázka**. Díky nastavené roli „právní analytik“ ví model, jakou tonalitu a styl použít (formální, věcný). Formát IRAC mu říká, jak strukturovat odpověď. *Není třeba dávat příklad hotového řešení – model formát dodrží. Všechny klíčové informace (termín dodání, množství, částka) jsou zahrnuty, takže model nemusí nic domýšlet. O1 pak v odpovědi interně projde logické kroky – identifikuje breach (porušení povinnosti dodat), aplikuje pravidlo (smluvní závazek), vyhodnotí, že A nesplnil, takže B má nárok atd., a sepiše to přehledně podle IRAC. Kdybychom stejnou věc chtěli po GPT-4, museli bychom ho nabádat např. „proved' detailní právní rozbor krok za krokem...“, jinak by nemusel uvést všechny úvahy. (Prompt Engineering for OpenAI's O1 and O3-mini Reasoning Models | Microsoft Community Hub) (Prompt Engineering for OpenAI's O1 and O3-mini Reasoning Models | Microsoft Community Hub)*

Příklad 2 – Matematická úloha:

Klasický GPT-4 často tápe u složitých matematických úloh, pokud mu explicitně nepřikážete „počítej krokově“. Reasoning model to udělá automaticky. **Prompt:**

„Vyřeš rovnici $x^3 - 2x^2 - 5x + 6 = 0$. Najdi všechny reálné kořeny a vysvětli postup.“

U GPT-4 by možná pomohlo dodat „step by step“, ale u reasoning modelu to není třeba – začne zkoušet dělitelnost polynomů, najde kořen $x=1$, provede polynomial division atd., to vše interně, a pak odpoví: kořeny jsou 1, -2, 3 a popíše, jak na to přišel. Uživatel nemusel model vést – ten díky RL tréninku ví „jak řešit“, nejen „co řešit“ ([A Visual Guide to Reasoning LLMs - by Maarten Grootendorst](#)).

Rekapitulace: Promptování reasoning modelů staví na obecných základech (jasné instrukce, kontext, formát) s tím rozdílem, že **model nepotřebuje tolik vodění za ruku ohledně logiky**. Nechte ho pracovat – stačí mu precizně říct, *co má vyřešit*, a dát mu k tomu případná data. Nemusíte (a neměli byste) ho zahlcovat příklady či pobízet k přemýšlení nahlas. Důležité je poskytnout mu všechny dílky skládačky (fakta, definice) potřebné k řešení a definovat, jak má výsledek vypadat. Pak můžete očekávat dobře strukturovanou, logicky zdůvodněnou odpověď. Samozřejmě stále platí modelářské „důvěřuj, ale prověřuj“ – u kritických výstupů raději proveďte ověření (ať už dalším dotazem nebo vlastní kontrolou).

3. Deep Research nástroje

S klasickými LLM (včetně reasoning modelů) narážíme na limit jejich znalostí: model neví nic o dění po datu svého tréninku a může si *vymýšlet* fakta (*halucinate*). **Deep Research** nástroje vznikly jako řešení tohoto problému. Jde o AI agenty, které propojují LLM s aktivním vyhledáváním informací online a jejich analýzou. Je to jako byste model zkombinovali s webovým vyhledávačem a rešeršistou v jednom – model sám prochází internet, sbírá relevantní data z více zdrojů a skládá z nich podrobnou odpověď s citacemi ([Deep Research in AI: Advanced Methodologies & Breakthrough Applications Explained](#)) ([New OpenAI 'Deep Research' Agent Turns ChatGPT into a Research Analyst -- Campus Technology](#)). Tyto nástroje dokáží během několika minut zpracovat to, co by lidskému analytikovi trvalo hodiny či dny: prohledají stovky stránek, vyfiltrují podstatné informace, porovnají je a sepišou strukturovanou zprávu s odkazy na původní zdroje ([Deep Research in AI: Advanced Methodologies & Breakthrough Applications Explained](#)) ([Deep Research in AI: Advanced Methodologies & Breakthrough Applications Explained](#)).

Charakteristika a princip: Deep Research (zkráceně DR) funguje jako autonomní agent nad LLM. Uživatel zadá komplexní dotaz nebo téma, a agent pak **iterativně vyhledává** na webu relevantní obsah ([Deep Research in AI: Advanced Methodologies & Breakthrough Applications Explained](#)) ([Deep Research in AI: Advanced Methodologies & Breakthrough Applications Explained](#)). Nejde o jednu prostou webovou query – agent provede **mnoho dotazů**: vždy si něco přečte, zjistí nové indicie a na jejich základě hledá dál, podobně jako by to dělal člověk postupně proklikávající odkazy ([Perplexity Deep Research Takes on OpenAI & Gemini](#)) ([Gemini: Try Deep Research and Gemini 2.0 Flash Experimental](#)). Během tohoto procesu využívá LLM nejen k generování konečné odpovědi, ale i k **rozboru nalezených textů** (dokáže „číst“ články, PDF dokumenty, dokonce interpretovat obrázky či grafy) ([Deep Research in AI: Advanced Methodologies & Breakthrough Applications Explained](#)). Důležité je, že DR agent informace **ověřuje křížově** – porovnává různé zdroje, hledá shody a rozpory ([Deep Research in AI: Advanced](#)

Methodologies & Breakthrough Applications Explained) (Deep Research in AI: Advanced Methodologies & Breakthrough Applications Explained). Tím výrazně snižuje riziko halucinací: nespolehá jen na „paměť“ LLM, ale **ověřuje fakta v reálných zdrojích** (Deep Research in AI: Advanced Methodologies & Breakthrough Applications Explained). Výstupem je pak **ucelená zpráva** – nejen stručná odpověď, ale celé shrnutí zjištěných poznatků, často členěné do přehledných sekcí, doplněné seznamem odkazů (citací) na původní články/studie (New OpenAI 'Deep Research' Agent Turns ChatGPT into a Research Analyst -- Campus Technology). OpenAI přirovnává odpověď DR agenta spíše k elaborátu zkušeného analytika než k běžné krátké odpovědi chatbota (Deep Research in AI: Advanced Methodologies & Breakthrough Applications Explained).

Konkrétní implementace se liší (viz dále dostupné nástroje), ale obecně DR agenti zahrnují tyto **klíčové komponenty**:

- **Multi-source analysis:** Paralelní i postupné čerpání z mnoha zdrojů a jejich porovnání (Deep Research in AI: Advanced Methodologies & Breakthrough Applications Explained). Agent například najde 10 článků k tématu, vyextrahuje z nich důležitá data a sleduje, zda se v nich opakují klíčové informace či zda si protiřečí.
- **Summarization:** Po nasbírání dat model provede jejich shrnutí do srozumitelné formy (Deep Research in AI: Advanced Methodologies & Breakthrough Applications Explained) (Deep Research in AI: Advanced Methodologies & Breakthrough Applications Explained). Komplexní témata vysvětlí jednoduše, z velkého objemu dat vypichne to podstatné. Často rozdělí výstup do kapitol s nadpisy, použije odrážky, tabulky apod. pro zlepšení čitelnosti.
- **Insight extraction:** Nejde jen o převyprávění zdrojů – agent aktivně **vytahuje závěry** a souvislosti (Deep Research in AI: Advanced Methodologies & Breakthrough Applications Explained). Identifikuje třeba trendy napříč studiemi, upozorní na klíčové statistiky, spojí informace dohromady a na konci může přidat syntézu či doporučení. V podstatě napodobí expertní zprávu „co z toho plyne“.
- **Citations a transparentnost:** Každé tvrzení v reportu je podloženo odkazem na zdroj (New OpenAI 'Deep Research' Agent Turns ChatGPT into a Research Analyst -- Campus Technology). Uživatel si tak může kliknout a ověřit, že informace opravdu pochází z důvěryhodného článku. DR agenti mívají i log, co při hledání dělají – některé (např. OpenAI) zobrazují průběh analýzy, takže vidíte, jaké dotazy a stránky procházejí (Perplexity Unveils Deep Research: AI-Powered Tool for Advanced Analysis - InfoQ) (Perplexity Unveils Deep Research: AI-Powered Tool for Advanced Analysis - InfoQ).

Celý proces běží několik minut autonomně. Například OpenAI Deep Research agent pracuje obvykle **5–30 minut** na jedné komplexní otázce (New OpenAI 'Deep Research' Agent Turns ChatGPT into a Research Analyst -- Campus Technology) (podle nastavení; uživatel mezitím čeká, jako by za něj pracoval výzkumník). Perplexity AI uvádí délku **pod 3 minuty** pro většinu dotazů díky optimalizacím (Perplexity Deep Research Takes on OpenAI & Gemini) (Perplexity Deep Research Takes on OpenAI & Gemini). Google Gemini agent to zvládá zhruba do **15 minut** (Perplexity Deep Research Takes on OpenAI & Gemini). Obecně platí, že čím důkladnější a přesnější výstup, tím déle agent hledá (např. OpenAI agent jde více do hloubky než rychlý Perplexity, viz srovnání níže).

Příklady praktického využití: Deep Research nástroje se uplatní tam, kde je třeba rychle získat detailní přehled o nějakém tématu či problému na základě aktuálních informací. Několik ilustrací reálných použití:

- **Finanční analýza:** Analytik na finančních trzích může agentovi zadat, aby zpracoval komplexní report o určitém odvětví – např. „Zjisti aktuální trendy v cenách komodit X, včetně faktorů ovlivňujících růst či pokles, a cituj analýzy z posledních 3 měsíců“. Agent projde finanční zpravodajství, blogy ekonomů, případně odborné analýzy a dodá souhrn s grafy a odkazy. OpenAI cílí právě na profesionály ve financích či politice, kterým DR ušetří hodiny práce při vyhledávání podkladů (New OpenAI 'Deep Research' Agent Turns ChatGPT into a Research Analyst -- Campus Technology) (New OpenAI 'Deep Research' Agent Turns ChatGPT into a Research Analyst -- Campus Technology).
- **Technologický výzkum:** Představme si postgraduální studentku, která se potřebuje rychle zorientovat v nejnovějších trendech senzorů pro autonomní vozidla kvůli prezentaci (Gemini: Try Deep Research and Gemini 2.0 Flash Experimental). Místo pročitání desítek článků zadá DR agentovi úlohu: „Zjisti, jaké typy senzorů se používají v autonomních autech, srovnej jejich výhody/nevýhody a popiš nové technologie na obzoru.“ Agent během pár minut vyhledá relevantní akademické papers, blogy z oboru, zprávy o produktech a dodá přehled, kde budou sekce o LiDARu, radarových a kamerových systémech, tabulka porovnávající dosah a přesnost, a zmínka o experimentálních senzorech s odkazy na články. To, co by jinak trvalo týden rešerší, má studentka za odpoledne.
- **Marketing a průzkum trhu:** Manažer připravující strategii může využít DR pro analýzu konkurence. Např.: „Zpracuj mi konkurenční analýzu pro trh chytrých hodinek – jaké hlavní modely uvedla konkurence za poslední rok, jejich klíčové funkce a cenová strategie, plus recenze uživatelů.“ Agent projde technologické weby, uživatelské recenze a fóra, a dodá zprávu s přehledem konkurentů, tabulkou parametrů hodinek a třeba i doporučením, kde je mezera na trhu. Google uvádí přesně takový use-case: malý podnikatel může DR využít k rychlému zmapování konkurence a návrhům, kam směřovat svůj byznys (Gemini: Try Deep Research and Gemini 2.0 Flash Experimental) (Gemini: Try Deep Research and Gemini 2.0 Flash Experimental).
- **Osobní rozhodování a consumer research:** I jednotlivci mohou DR využít pro kvalifikovaná rozhodnutí. Například někdo chce koupit nové auto – místo pročitání recenzí různě po webu zadá: „Porovnej mi modely aut A, B, C z hlediska spolehlivosti, spotřeby a bezpečnosti. Zjisti z testů a recenzí, jak si vedou, a doporuč, které vybrat.“ Agent posbírá údaje z auto-magazínů, výsledky nárazových testů, zkušenosti uživatelů na fórech a předloží přehled s tabulkou plus odkaz na zdroje. OpenAI zmiňuje, že DR může asistovat i u spotřebitelských rozhodnutí, jako je porovnání nábytku či elektroniky (New OpenAI 'Deep Research' Agent Turns ChatGPT into a Research Analyst -- Campus Technology). Rozdíl oproti běžnému vyhledávání je v tom, že dostanete **syntézu**, ne jen seznam odkazů.

Zkrátka, **Deep Research** je užitečný všude tam, kde byste normálně strávili spoustu času hledáním aktuálních informací a jejich čtením. DR agent to udělá za vás a naservíruje hotovou zprávu. Samozřejmě má smysl u složitějších dotazů – ptát se ho na datum narození slavné osoby je zbytečné (to zvládne Google okamžitě). Je určen pro komplexní výzkumné dotazy, „**komplexní otázky vyžadující expertní úroveň analýzy**“ ([New OpenAI 'Deep Research' Agent Turns ChatGPT into a Research Analyst -- Campus Technology](#)).

Dostupné DR nástroje: Aktuálně (začátek 2025) existuje několik hlavních nástrojů a služeb, které funkcionalitu Deep Research nabízí. Následující tabulka shrnuje jejich výrobce, vlastnosti, omezení a cenu:

Nástroj (výrobce)	Funkce a vlastnosti	Omezení	Cena a dostupnost
OpenAI ChatGPT – Deep Research (OpenAI)	Pokročilý DR agent integrovaný v ChatGPT. Běží na specializované verzi modelu GPT-4 (tzv. o3) s důrazem na důkladné ověřování a analýzu (New OpenAI 'Deep Research' Agent Turns ChatGPT into a Research Analyst -- Campus Technology). Stráví až ~30 min. samostatným hledáním, dokáže zpracovat texty, obrázky i PDF (Deep Research in AI: Advanced Methodologies & Breakthrough Applications Explained) (Deep Research in AI: Advanced Methodologies & Breakthrough Applications Explained). Výsledkem je velmi detailní report (často desítky stránek) se strukturovanými sekcemi, tabulkami, grafy a ~desítkami citací (Perplexity Unveils Deep Research: AI-Powered Tool for Advanced Analysis - InfoQ) (Perplexity Unveils Deep Research: AI-Powered Tool for Advanced Analysis - InfoQ). Přesnost má v testech nejvyšší – dosáhl 26,6 % v Humanity's Last Exam (vs. GPT-4o jen 3,3 %) (New OpenAI 'Deep Research' Agent Turns ChatGPT into a Research Analyst -- Campus Technology), poráží i konkurenci (Gemini ~6 %) díky iterativnímu přístupu a minimálním halucinacím.	Velmi vysoké nároky na čas a zdroje – odpověď trvá 5–30 min. a je „drahá“. Aktuálně omezen na 100 dotazů měsíčně a jen pro uživatele v USA (New OpenAI 'Deep Research' Agent Turns ChatGPT into a Research Analyst - Campus Technology). Není zatím globálně přístupný (stav únor 2025). Chybí mu <i>real-time</i> multimodální funkce v průběhu (grafy generuje až na závěr). Je to prémiová funkce – není součástí běžného Plus účtu.	\$200/měsíc – dostupné v rámci ChatGPT Pro (samostatná nejvyšší placená vrstva) (New OpenAI 'Deep Research' Agent Turns ChatGPT into a Research Analyst - Campus Technology). Postupně má být zpřístupněn i uživatelům ChatGPT Plus/Team/Enterprise (s nižšími limity) (New OpenAI 'Deep Research' Agent Turns ChatGPT into a Research Analyst - Campus Technology).
Perplexity AI – Deep Research (Perplexity)	Samostatný webový nástroj a API zaměřený na rychlý DR. Prohledá desítky zdrojů a do ~3 minut vygeneruje report (Perplexity Deep Research Takes on OpenAI & Gemini) (Perplexity Deep Research Takes on OpenAI & Gemini). Využívá open-source reasoning model (DeepSeek-R1) a optimalizovaný postup, takže je velmi rychlý (v testu byl ~9× rychlejší než ChatGPT DR) (Perplexity Unveils Deep Research: AI-Powered Tool for Advanced Analysis - InfoQ). Přesnost překvapivě vysoká – ~21 % v HLE benchmarku, blízko OpenAI (díky využití DeepSeek) (Perplexity Unveils Deep Research: AI-Powered Tool for Advanced Analysis - InfoQ). Výstup obsahuje dobře strukturovaný text, tabulky a citace ~50 zdrojů (Perplexity Unveils Deep Research: AI-Powered Tool for Advanced Analysis - InfoQ), lze ho exportovat do PDF. Velké plus: zcela zdarma v základní verzi.	O něco menší hloubka analýzy – při rychlosti ❤️ min občas nestihne dohledat úplně vše (Perplexity Deep Research Takes on OpenAI & Gemini) (Perplexity Deep Research Takes on OpenAI & Gemini). V porovnání s OpenAI DR občas chybí některá data nebo detailnější rozbor, když nejsou snadno k nalezení (Perplexity Deep Research Takes on OpenAI & Gemini). Taktéž vizuální výstup je prostší – jen statický text a tabulky, žádné interaktivní grafy (Perplexity Deep Research Takes on OpenAI & Gemini). Omezení: anonymní uživatel má limit 5 dotazů denně (Perplexity Deep Research Takes on OpenAI & Gemini); pro vyšší limity nutno předplatit Pro.	Zdarma (5 dotazů/den pro neregistrované). \$20/měsíc Pro verze nabízí až 500 dotazů denně a rychlejší odezvy (Perplexity Deep Research Takes on OpenAI & Gemini). Dostupné globálně přes webový prohlížeč, mobilní app (iOS, Android) i macOS app (Perplexity Deep Research Takes on OpenAI & Gemini).

Nástroj (výrobce)	Funkce a vlastnosti	Omezení	Cena a dostupnost
Google Gemini – Deep Research (Google)	DR agent integrovaný do AI asistenta Gemini (nástupce BARDu). Při dotazu v režimu “Deep Research” nejprve vygeneruje plán výzkumu, který dá uživateli ke schválení/upravení (Gemini: Try Deep Research and Gemini 2.0 Flash Experimental). Poté několik minut prohledává Google a postupně zpřesňuje analýzu (Gemini: Try Deep Research and Gemini 2.0 Flash Experimental). Využívá vlastní LLM Gemini 1.5 (resp. 2.0 Flash) a Google Search (má přímý přístup do indexu). Kontext okna až 1 milion tokenů umožňuje pojmout opravdu rozsáhlé podklady (Gemini: Try Deep Research and Gemini 2.0 Flash Experimental). Výsledná zpráva je exportovatelná do Google Docs a obsahuje odkazy. Výhodou je silná integrace s Google ekosystémem – umí využít znalosti o vyhledávání a brzy možná naváže i na Workspace (GDrive data apod.).	Pouze v angličtině a v rámci omezené služby Gemini Advanced (dostupné zatím v několika zemích). Výsledky v testech přesnosti zatím zaostávaly – Gemini Thinking (starší verze) dala jen ~6,2 % v HLE (Perplexity Deep Research Takes on OpenAI & Gemini). Plánování výzkumu je fajn, ale bylo zaznamenáno, že může uvíznout ve statickém plánu (nehledá mimo předem schválené body) (Perplexity Deep Research Takes on OpenAI & Gemini). Interaktivita je menší než u OpenAI – uživatel vstupuje jen na začátku (schválení plánu), pak už agent jede sám. Cena je nižší, ale výkon a hloubka analýzy také.	\$20/měsíc v rámci tarifu Gemini Advanced (případně součást předplatného Google One v nejvyšší úrovni). Uvedeno v prosinci 2024 (Perplexity Deep Research Takes on OpenAI & Gemini). Dostupnost: webová Gemini aplikace (zatím globálně, jazyk EN), mobilní Gemini app (brzy), integrace do Google Workspace se chystá (Gemini: Try Deep Research and Gemini 2.0 Flash Experimental).
Další (ostatní)	Bing Chat (Microsoft): Chatbot od Microsoftu také nabízí mód „Výzkum“ – využívá model GPT-4 s přímým Bing vyhledáváním. Umí odpovídat s citacemi, ale hloubka je menší (spíše odpověď na jeden dotaz s citacemi, ne dlouhý report). YouChat, Phind aj.: Podobně kombinují LLM s webem; Phind se specializuje na vývojářské dotazy s kódem. xAI Grok: Model od Muska s internetovým přístupem (přes X/Twitter), ale zatím není robustní generální řešeršní nástroj.	Bing Chat: limitován na relace po pár dotazech, někdy odmítá příliš dlouhé výzkumy. Ostatní: různé kvality zdrojů, někdy citace jen na Wikipedii apod. Grok: omezen hlavně na data z X, ne plný web.	Bing: zdarma pro uživatele Edge/Bing (omezené počty dotazů denně). Phind/YouChat: zdarma online. Grok: beta přístup pro vybrané uživatele X Premium (stav 2024).

(Poznámka: Údaje o přesnosti (HLE benchmark) a časech vychází z veřejných testů a oznámení v únoru 2025 (New OpenAI ‘Deep Research’ Agent Turns ChatGPT into a Research Analyst -- Campus Technology) (Perplexity Deep Research Takes on OpenAI & Gemini). Míra přesnosti znamená schopnost zodpovědět expertní otázku napříč 100 tématy – HLE = Humanity’s Last Exam, viz výše.)

Z výše uvedeného je patrné, že OpenAI Deep Research nabízí nejhlubší analýzu (nejvíce zdrojů, interaktivní proces), ale je exkluzivní a drahý. Perplexity představuje skvělý kompromis – je překvapivě schopný a dostupný každému zdarma, byť někdy méně dopodrobna. Google Gemini DR je někde mezi: systematický, ale zatím ne tak přesný ani dostupný všem. Konkurence se rychle vyvíjí a můžeme očekávat, že brzy se přidají další platformy a zlepší i ty stávající.

Nejlepší praktiky promptingu DR nástrojů: Aby AI agent odvedl dobrou práci, musíme mu správně formulovat zadání – trochu jinak než u běžného LLM. DR prompt je spíš výzkumná otázka nebo zadání projektu než jednoduchý dotaz. Zde jsou doporučení:

- Jasně definujte cíl výzkumu:** Uveďte, co přesně chcete zjistit. Místo vágního „Řekni mi něco o elektrických autech“ napište raději „Zjistí aktuální statistiky prodeje elektrických aut v Evropě za posledních 5 let a identifikuj hlavní faktory, které ovlivnily nárůst či pokles v jednotlivých letech.“ Čím konkrétnější otázku položíte, tím lépe agent ví, na co se při hledání zaměřit.
- Rozdělte komplexní dotaz na podotázky:** Můžete v promptu vyjmenovat několik bodů, které chcete pokrýt. Např.: „....Zpráva by měla obsahovat: (1) statistiky adopce iOS vs Android za posledních 5 let v top 10 vyspělých a 10 rozvojových zemích; (2) procento populace, které se učí druhý jazyk v těchto zemích; (3) trendy v penetraci mobilů; (4) doporučení, na které trhy cílit novou aplikaci pro překlad.“ Tím v podstatě agentovi **nadiktujete osnovu**. V testech se ukázalo, že takový strukturovaný dotaz vede k velmi kvalitní odpovědi – agent pokryje všechny body. (Ukázkový prompt tohoto typu viz níže.) (Perplexity Deep Research Takes on OpenAI & Gemini)
- Specifikujte formu výstupu:** Podobně jako u LLM, i zde můžete říct, jak má výstup vypadat. DR agenti sice mají svůj výchozí formát (hlavička, přehled, kapitoly, zdroje), ale lze je trochu ovlivnit. Např. „....uved výsledky v přehledné tabulce a připoj doporučení na závěr.“ Perplexity i OpenAI v ukázkách reagují na požadavek zahrnout tabulky či přehledy (Perplexity Deep Research Takes on OpenAI & Gemini). Pokud něco takového chcete, řekněte to rovnou.

- **Doplňující kontext (volitelně):** Na rozdíl od reasoning modelu nemusíte DR agentovi dodávat znalosti – on si je najde sám. Ale pokud třeba **víte o konkrétním zdroji, který nechcete vynechat**, můžete ho zmínit. Např. „...zaměř se hlavně na data od WHO a Světové banky.“ Agent pak upřednostní tyto zdroje ve výsledku, pokud existují. Stejně tak lze specifikovat *typ* zdrojů: „...použij především akademické studie.“ (Nástroj pak bude hledat spíše na Google Scholar než na populárních webech.) Některé DR mají i volby v UI k preferencím zdrojů.
- **Zkontrolujte a upravte plán (pokud to nástroj umožňuje):** Google Gemini před zahájením hledání vygeneruje plán – např. jaké podotázky zodpoví ([Gemini: Try Deep Research and Gemini 2.0 Flash Experimental](#)). Využijte toho a doplňte či opravte ho, než spustíte samotný výzkum. Když vidíte, že agent na něco zapomněl, přidejte to. U jiných agentů takovou možnost nemáte, ale můžete podobný „plán“ rovnou sami načrtnout v promptu (viz výše).
- **Trpělivost během běhu:** Některé agenty (OpenAI) zobrazují průběh analýzy – např. kolik procent už prohledali. Nezasahujte do toho, nechte je pracovat. U OpenAI běží i 10+ minut, což je normální ([Perplexity Deep Research Takes on OpenAI & Gemini](#)). Přerušit byste to měli jen, pokud výzkumná otázka ztratila smysl.
- **Projděte si citace a závěry:** Po vygenerování výstupu doporučujeme **prolínout zdroje**, které agent uvedl, zvláště pokud na základě toho máte dělat důležitá rozhodnutí. DR značně snižuje riziko halucinací, ale stále se může stát, že některý detail není přesný nebo interpretace je diskutabilní ([New OpenAI 'Deep Research' Agent Turns ChatGPT into a Research Analyst -- Campus Technology](#)). Rychlá kontrola klíčových zdrojů vám dá jistotu. (Odborník z University of Surrey varoval, že i když AI udělá analýzu, člověk musí věnovat čas ověření – jinak může přehlédnout, pokud AI něco špatně vyhodnotila ([New OpenAI 'Deep Research' Agent Turns ChatGPT into a Research Analyst -- Campus Technology](#)).)
- **Pokračující dotazy (follow-up):** Stejně jako u chatu můžete po obdržení reportu klást doplňující otázky. Např.: „Výborně, a teď zjisti totéž pro rok 2023 zvláště pro ČR.“ Agent pak může navázat a doplnit, aniž by musel znovu zpracovávat vše od začátku. OpenAI i Google podporují takové navazující dotazy – model má interně uložený kontext, co zjistil, a reaguje rychleji na rozšíření dotazu.

Ukázkový prompt pro Deep Research: (inspirováno skutečným srovnávacím testem)

„Pomoz mi zjistit **míru adopce iOS a Androidu**, procento populace, která se chce učít další jazyk, a **změnu mobilní penetrace** za posledních 5 let u **10 nejvyspělejších a 10 rozvojových zemí** (podle HDP). Seřaď výsledky do přehledné **tabulky** a připoj doporučení, na **které trhy** by se měla zaměřit nová aplikace na výuku jazyků.“

Tento dotaz má jasné sub-položky (adopce OS, zájem o jazyky, penetrace mobilů) a vymezené skupiny zemí. Agent tedy nejspíš:

1. Najde seznam top 10 developed a developing zemí dle HDP.
2. Vyhledá statistiky OS (iOS vs Android market share) v těchto zemích za posledních 5 let.
3. Vyhledá průzkumy či data o tom, kolik % lidí se učí druhý jazyk.
4. Najde data o mobilní penetraci (uživatelé mobilů v populaci).
5. Vše ztabulkuje a vyvodí, které země jsou perspektivní pro app na učení jazyků (např. vysoká penetrace mobilů, ale nízké procento lidí učících se jazyk => potenciál růstu).

Ve skutečném testu na podobný dotaz: **Perplexity** to zvládl za ~3 minuty, ale některé metriky neměl kompletní (např. chyběla úplná data pro všechny země), **Google Gemini** to vyřešil za ~6 min se solidními tabulkami a rozbořem, a **OpenAI DR** agent strávil ~11 minut a dodal extrémně podrobnou analýzu včetně trendů a grafů ([Perplexity Deep Research Takes on OpenAI & Gemini](#)) ([Perplexity Deep Research Takes on OpenAI & Gemini](#)). Otevřená ukázka zmiňuje, že ChatGPT Deep Research z tohoto promptu udělal 41stránkový report se 75 citacemi ([Google Deep Research vs. OpenAI Deep Research: The Future of ...](#)) – skutečně analytická práce. Taková ilustrace ukazuje rozdíl v hloubce: Perplexity dal rychlé odpovědi, OpenAI šel **mnohem více do hloubky** (včetně vizualizací a strategických doporučení) ([Perplexity Deep Research Takes on OpenAI & Gemini](#)) ([Perplexity Deep Research Takes on OpenAI & Gemini](#)). Volba nástroje tedy záleží na tom, zda preferujete rychlost, nebo maximální detail.

Rekapitulace: Deep Research nástroje posouvají schopnosti AI z „odpovídání na znalostní otázky“ k **provádění plnohodnotného výzkumu**. Využívají LLM k analýze informací z webu, čímž **eliminují omezení tréninkových dat** (dokáží pracovat s nejnovějšími poznatky) a výrazně **omezují halucinace** (všechna fakta si ověřují ve zdrojích). Na reálných příkladech jsme viděli, že DR umí generovat komplexní reporty pro odborné i praktické účely – od vědeckých rešerší přes obchodní analýzy až po osobní rozhodování. Při používání je klíčové správně formulovat dotaz jako výzkumné zadání, případně strukturovat požadavky. Současné hlavní nástroje jsou OpenAI (nejdůkladnější, ale drahý), Perplexity (dostupný a rychlý) a Google Gemini (integrace s Google ekosystémem). Všude platí: AI výrazně zrychlí sběr a shrnutí informací, **ale člověk by měl finální report vždy projít a kriticky zhodnotit**, zejména u závažných rozhodnutí.

4. Prompting Deep Research nástrojů

Nyní se zaměříme na to, jak správně *promptovat* (zadávat dotazy) Deep Research agentům a v čem se to liší od promptování reasoning modelů či běžných LLM. V předchozí kapitole jsme již naznačili nejlepší praktiky pro DR – zde to shrneme a porovnáme.

Prompting reasoning vs. DR modelů – rozdíly: U reasoning modelu (bez připojení k webu) je často potřeba dodat modelu *všechny informace* k řešení v promptu, protože sám nic nedohledá ([Prompt Engineering for OpenAI's O1 and O3-mini Reasoning Models | Microsoft Community Hub](#)). Naproti tomu DR agent si informace aktivně opatří – stačí mu říct, co *zjistit*, nemusíte mu data v promptu vypisovat. To ovšem znamená, že dotaz pro DR může být pojat obecněji (např. „Zjistí nejnovější vývoj v oboru XYZ“), zatímco reasoning model by na takový dotaz odpovídal jen z paměti (a mohl by halucinovat nebo neznat nové údaje). **U DR se tedy nebojte pokládat i velmi široké otázky**, které vyžadují prohledat mnoho zdrojů – od toho tam agent je.

Další rozdíl je v tom, že reasoning model pracuje přímo v konverzaci – můžete ho vést krok za krokem (např. postupně mu dávat data k analýze). Naproti tomu DR agent pracuje spíše *batchově* – dáte mu komplexní zadání a on se na delší dobu „odmlčí“ a pak přinese výsledek.

Prompt DR nástroje je tedy často delší a komplexnější než typický prompt do chatu. Můžete do něj zahrnout více sub-otázek, požadavků na výstup atd., protože nečekáte okamžitou jednovětnou odpověď, ale strukturovanou zprávu.

U reasoning modelu jsme varovali nepoužívat „*step-by-step*“ instrukce, protože model to dělá sám interně. Podobně u DR agentů nemusíte (a neměli byste) explicitně popisovat, *jak* mají hledat – např. neříkejte „*nejdřív vyhledej tohle, pak tamto*“. Tyto agenty mají vlastní heuristiku vyhledávání a váš zásah by mohl spíše omezit jejich efektivitu. Místo toho formulujte *co zjistit*. Výjimkou je Google Gemini, kde můžete upravit navržený plán – to je ale součást UX, nikoli promptu jako takového. V promptu pro DR tedy **popisujete cíl a obsah výstupu, ne proces** (ten nechte na agentovi).

Klíčové principy promptingu DR nástrojů:

- Výzkumná otázka místo přímého dotazu:** Formulujte vstup spíše jako zadání seminární práce nebo reportu. Např. „*Analyzuj dopady technologie X na odvětví Y za poslední dekádu, zahrň hlavní milníky a současný stav výzkumu.*“ Tak agent ví, že má hledat historický vývoj, klíčové studie, aktuální review atd. Kdybyste totéž dali klasickému LLM, buď by to nevěděl, nebo by halucoval – DR agent to zjistí.
- Nevyhledávejte manuálně předem:** Někteří uživatelé mají tendenci do promptu cpát odkazy a citace, které sami našli. To není potřeba a může to být spíš na škodu (agent by pak ty odkazy zvýhodnil a jiné možná ignoroval). Lepší je uvést *druhy zdrojů* („... použij oficiální statistiky WHO...“). Výjimka je, pokud máte konkrétní dokument, který *určitě* má být zahrnut (třeba interní PDF zprávu) – ten můžete nahrát (u OpenAI DR) nebo dát URL (některé agenty umí číst z URL). Jinak ale platí: **svěřte vyhledávání agentovi**, je v tom efektivnější.
- Kontrola a iterace skrze follow-up místo jednoho promptu:** Pokud výsledek není úplně dle očekávání, je často lepší položit upřesňující dotaz *navazující* než hned celý prompt formulovat znovu. Agent už má totiž spoustu dat z prvního běhu. Např. když Perplexity něco opomene, stačí se doptat: „*Opomněl jsi zmínit XYZ, doplň to prosím.*“ a on rychle dohledá jen tu chybějící část. V promptu byste to příště zmínili už od začátku. Čili prvotní prompt zkuste co nejlépe, ale nebojte se výsledek **interaktivně doleštit dalším dotazem** – je to rychlejší než vše formulovat nanovo.
- Specifika různých nástrojů – modely v pozadí:** Je užitečné vědět, *jaký LLM daný DR nástroj pohání*, protože to ovlivní styl a schopnosti:
 - OpenAI Deep Research používá variantu GPT-4 (o3) s **důrazem na reasoning** ([New OpenAI 'Deep Research' Agent Turns ChatGPT into a Research Analyst -- Campus Technology](#)). To znamená, že odpovědi jsou velmi detailní, košaté a prakticky vždy správně podložené. Styl je spíše formální a odborný. Také to znamená, že nepotřebuje žádné navádění k „přemýšlení“ – prompt může být stručnější na instrukce k řešení, protože GPT o3 si samo zvolí postup. Klidně i obecnější dotazy zvládne bravurně, ale počítejte s dlouhým čekáním.
 - Perplexity Deep Research jede na open-source modelu (**DeepSeek-R1**) ([Perplexity Unveils Deep Research: AI-Powered Tool for Advanced Analysis - InfoQ](#)). Ten má vysoký výkon v logických úlohách, ale přeci jen není tak vyladěný na výstupní stylistiku jako GPT-4. Odpovědi mohou být stručnější a někdy méně „vyleštěné“. Můžete proto do promptu klást větší důraz na formu (např. „*udělej to srozumitelně i pro laiky*“ nebo „*vyvaruj se příliš technického detailu*“), pokud je to potřeba. DeepSeek je velmi rychlý, takže Perplexity si může dovolit krátké iterace hledání – prompt ale stejně napište kompletní, nečekejte, že se agent *doptá*. (Perplexity se na rozdíl od GPT-4 DR uživatele na nic během běhu nezeptá.)
 - Google Gemini DR používá model **Gemini 1.5/2.0**. Ten zatím nemá takovou logickou hloubku jako o1 nebo DeepSeek, proto je vhodné prompt formulovat jednoznačně a **rozsáhleji uvést kontext**. Co je u OpenAI samozřejmost (že to najde relevantní data), může u Gemini vyžadovat popostrčení. Například při dotazu na lékařský výzkum bychom v promptu mohli uvést „... *zaměř se na recenzované studie a metaanalýzy z posledních 5 let.*“ Tím Gemini nasměrujeme k lepším zdrojům (jinak by mohl brát i popularizační weby, kde hrozí nižší kvalita). Znalost, že Gemini má 1M tokenů kontext, můžeme využít: klidně do promptu vložte větší text, co už máte, k analýze – on to unese.
- Tón a jazyk:** Většina DR agentů momentálně funguje nejlépe v angličtině. Pokud pokládáte dotaz v jiném jazyce, může se stát, že agent stejně bude hledat primárně anglické zdroje a odpověď pak přeloží (jako to dělá Bing). Kvalita překladu i nalezených zdrojů nemusí být ideální. Obecně pro co nejlepší výsledky se doporučuje **prompt psát anglicky**, pokud je to možné, protože tak využijete plný potenciál nástroje. (Samozřejmě finální zprávu si pak můžete nechat přeložit třeba samotným ChatGPT.) To je rozdíl oproti reasoning modelům typu o1 – ty jsou trénované vícejazyčně a poradí si i s češtinou slušně. U DR je ale primární doména angličtina, jelikož většina online obsahu a modelů (Bing, Google) je optimalizovaná pro EN.

Přehled modelů v pozadí DR nástrojů: Jak bylo řečeno, OpenAI vs. Perplexity vs. Gemini používají různé LLM, což se projevuje:

- **OpenAI DR (GPT-4 o3)** – nejpokročilejší model, který **tráví více času myšlením** a hledáním, než odpoví ([Introducing OpenAI o1 | OpenAI](#)) ([New OpenAI 'Deep Research' Agent Turns ChatGPT into a Research Analyst -- Campus Technology](#)). Je to vidět na kvalitě – v testu HLE získal 26,6 %, což je o *řád lepší* než klasický GPT-4o (3,3 %) ([New OpenAI 'Deep Research' Agent Turns ChatGPT into a Research Analyst -- Campus Technology](#)). Prakticky to znamená, že tam kde běžný ChatGPT tápe, DR agent uspěje. Tuto sílu dává RL trénink na chain-of-thought a možná i větší „**výpočetní rozpočet**“ **při inferenci** (OpenAI naznačilo, že o1/o3 mají možnost strávit více kroků výpočtu při hledání odpovědi než GPT-4o) ([o1: A Technical Primer - LessWrong](#)) ([OpenAI's o1 is a misunderstood model - by Charlie Guo](#)). Pro promptování to znamená: můžete si dovolit být ambiciózní – zadat velmi komplexní úkol – a model ho pravděpodobně zvládne do detailu. Jen pamatujte na limit 100 dotazů měsíčně, takže s tím neplýtvat na maličkosti.
- **Perplexity DR (DeepSeek)** – stojí na open-source **DeepSeek-R1** ([Perplexity Unveils Deep Research: AI-Powered Tool for Advanced Analysis - InfoQ](#)). Ten je výsledkem destilace a RL tréninku komunity a dosahuje výkonu srovnatelného s GPT-4 v mnoha benchmarcích ([DeepSeek](#)) ([DeepSeek](#)). Nicméně nemá tak širokou znalost a cit pro jazykové finesy. V praxi si s tím Perplexity poradí tak, že model pustí spíše „na krátko“, aby stihl do 3 min. vše projít, a soustředí se na fakta. Prompt by měl být u Perplexity velmi konkrétní ohledně toho, co chceme – protože jakmile něco explicitně zmíníte, agent to zahrne. Naopak co nezmíníte, může vynechat, pokud to sám nepovažuje za klíčové. Je tedy vhodné *vyjmenovat, co má report obsahovat* (jako v ukázkovém promptu výše). Tím minimalizujete šanci, že něco podstatného vynechá kvůli časovému limitu.
- **Gemini DR (Gemini LLM)** – Gemini je odlišný v tom, že silně využívá **Google search technologie** a agentní systém plánování ([Gemini: Try Deep Research and Gemini 2.0 Flash Experimental](#)) ([Gemini: Try Deep Research and Gemini 2.0 Flash Experimental](#)). Model tedy spoléhá, že mu Google dobře najde relevantní věci. Prompt by proto měl obsahovat klíčová *klíčová slova* (😊) podobně jako byste je zadali do Google. Např. dotaz „*How has CRISPR-Cas9 evolved in gene therapy research between 2018 and 2024, and what are the major breakthroughs?*“ – tučná slova jsou to, co určitě zahrnout. Kdybych to samé položil volně „*Summarize gene therapy advances recently*“, možná by hledal všelicos. Takže pro Gemini **vytipujte terms, co by měl hledat**. Jinak se model Gemini stále zlepšuje – verze 2.0 (experimentální) má výrazně vyšší kapacitu i rychlost ([Gemini: Try Deep Research and Gemini 2.0 Flash Experimental](#)) ([Gemini: Try Deep Research and Gemini 2.0 Flash Experimental](#)), takže brzy může konkurovat OpenAI. Prozatím je však vhodné prompt formulovat spíše detailně, nenechávat nic na inteligenci modelu (která je o něco slabší v reasoning než u OpenAI).

Shrme-li, **promptování DR agentů** je o *pečlivém nadefinování cíle a obsahu reportu*, zatímco samotný proces rešerše necháváme na AI. Rozdíly mezi nástroji jsou v nuance: OpenAI zvládne hodně i z kusého zadání (ale raději využijte tu možnost a popište požadavky důkladně, ať využije plný potenciál), Perplexity potřebuje jasné instrukce co pokrýt (aby nic důležitého nevynechal) a Google uvítá, když mu promptem „naservírujete“ relevantní pojmy pro vyhledávání. Ve všech případech platí, že dobře strukturovaný dotaz vede k lépe strukturované odpovědi. A naopak mlhavé zadání (třeba „*Udělej mi výzkum o medicíně*“) může vést k povrchnímu výsledku i od sebelepší AI.

Rekapitulace: Promptování Deep Research nástrojů spočívá v položení komplexní otázky, která jasně vymezuje *co má výstup obsahovat*. Na rozdíl od klasického promptu u chatbota zde můžeme (a měli bychom) zahrnout více detailů a požadavků, klidně formou odrážek. DR agent nepotřebuje, abychom mu říkali *jak* má postupovat – to je jeho úkol; potřebuje ale vědět, *co od něj chceme zjistit a v jaké formě*. Kvalitní prompt u DR tak vypadá spíše jako zadání úkolu pro lidského analytika. Různé nástroje (OpenAI, Perplexity, Gemini) mají v pozadí různé modely, ale pro uživatele je podstatné, že všem by měl dát dobré vstupní instrukce a pak nechat AI pracovat. Finální výstup je vhodné ověřit a případně dalším dotazem upřesnit či doplnit. Při dodržení těchto zásad dokáží DR agenti produkovat působivě kvalitní rešerše v rekordním čase.

5. Shrnutí

Klíčová fakta z celého materiálu:

- **Reasoning AI modely** představují novou generaci LLM, které mají schopnost vnitřně uvažovat ve více krocích. Od klasických modelů (ChatGPT-4 apod.) se liší tím, že netvoří odpověď jedním skokem, ale **rozkládají problém na dílčí úvahy** a ty postupně řeší ([A Visual Guide to Reasoning LLMs - by Maarten Grootendorst](#)). To jim umožňuje výrazně lepší výkon v komplexních úlohách (např. složité matematické a logické problémy) ([Introducing OpenAI o1 | OpenAI](#)). OpenAI uvedlo řadu o1 (2024) trénovanou RL, která v testech překonává GPT-4 o řád. Google odpovídá modelem Gemini s agentním plánováním. Existují i open-source varianty jako DeepSeek.
- **Promptování reasoning modelů** vyžaduje trochu jiný přístup než u klasických LLM. Základní pravidla (jasnost, kontext, požadovaný formát) platí, ale není potřeba modelu podsouvat řetězec myšlenek – **reasoning modely si „chain-of-thought“ generují samy** ([Prompt Engineering for OpenAI's O1 and O3-mini Reasoning Models | Microsoft Community Hub](#)). Naopak je důležité dodat případné informace, které model nemusí znát (aktuální či specifická data), protože nemá přístup k externím znalostem ([Prompt Engineering for OpenAI's O1 and O3-mini Reasoning Models | Microsoft Community Hub](#)). Nejlepší praxí je formulovat dotaz přímo a nechat model, ať si postup řešení rozvrhne. Při promptování o1/o3 se vyhněte zbytečným příkladům či krokovacím pokynům, spíše využijte možnosti nastavit roli a požadovaný výstup. Reasoning model typicky zvládne komplexní úlohu na první dobrou – pokud ne, iterujte promptem nebo se doptávejte pro ověření.

- **Deep Research (DR) nástroje** integrují LLM s webovým vyhledáváním a umožňují AI **autonomně provádět online rešerše**. Z vašeho dotazu vytvoří sérii webových dotazů, přečtou si desítky stránek, extrahují podstatné informace a sepišou z nich **ucelenou zprávu s citacemi** ([Deep Research in AI: Advanced Methodologies & Breakthrough Applications Explained](#)) ([Deep Research in AI: Advanced Methodologies & Breakthrough Applications Explained](#)). Tím překonávají omezení běžných LLM – umí pracovat s aktuálními informacemi a ověřovat fakta napříč zdroji, čímž drasticky snižují halucinace. Jsou ideální pro komplexní otázky, kde by člověk jinak musel dlouho vyhledávat (analýzy trendů, srovnání produktů, odborné přehledy). Hlavními představiteli jsou ChatGPT Deep Research (OpenAI), Perplexity AI a Google Gemini Deep Research. Liší se hloubkou, rychlostí a dostupností: OpenAI je nejdůkladnější (a nejdražší), Perplexity velmi rychlý a dostupný zdarma, Google propojený s vyhledávačem ale zatím méně přesný.
- **Promptování DR nástrojů** se podobá zadání úkolu pro analytika. Je potřeba jasně definovat, *co má výzkum zjistit*, a ideálně strukturovat dotaz do několika konkrétních požadavků (podotázek). Na rozdíl od chatbota zde nevedeme dialog krok za krokem (agent si postup řídí sám), takže do jednoho promptu dáme všechny instrukce. Nemusíme dodávat fakta (agent si je najde), ale můžeme specifikovat preferované zdroje či formu výstupu. Důležitá je *konkrétnost* – agent pak pokryje všechny uvedené body. Po vygenerování reportu lze klást navazující dotazy pro doplnění. Celkově: **čím lépe vymezíte zadání, tím kvalitnější a užitečnější bude výstup DR agenta**.

Nejdůležitější informace k zapamatování:

- *Reasoning LLMs vs standard LLMs*: Reasoning modely (o1, Gemini atd.) „**přemýšlejí**“ **před odpovědí** a zvládají tím pádem mnohem těžší úkoly s menší chybou (Introducing OpenAI o1 | OpenAI). Jsou však pomalejší a méně vhodné pro jednoduché dotazy – tam stačí klasický LLM.
- *Prompting reasoning models*: **Nenuťte je myslet krokově – dělají to samy** ([Prompt Engineering for OpenAI's O1 and O3-mini Reasoning Models | Microsoft Community Hub](#)). Soustřeďte se na přesné položení problému a poskytnutí relevantních informací. Uveďte, jaký typ odpovědi chcete (styl, formát). Příklad: místo „*Vypočti těžký příklad*“ dejte konkrétní příklad a požadavek „*vysvětli řešení*“ – reasoning model pak bezpečně dojde k správnému výsledku a popíše postup.
- *Deep Research agents*: Jsou to vaše **AI výzkumní asistenti** – umí za vás projít stovky webových zdrojů a sepsat fundovaný přehled s odkazy ([Deep Research in AI: Advanced Methodologies & Breakthrough Applications Explained](#)). Oproti běžnému vyhledávání ušetří obrovské množství času a dokáží integrovat poznatky z různých míst. Výstupy mají formu reportů (často několik stran textu). Důvěryhodnost je vyšší než u klasických LLM, protože vše citují a kontrolují ([New OpenAI 'Deep Research' Agent Turns ChatGPT into a Research Analyst -- Campus Technology](#)), ale i tak je vhodné výsledky ověřovat.
- *Prompting DR*: Napište **detailní zadání** toho, co chcete zjistit – klidně v bodech. Nebojte se ptát na více věcí najednou, agent je pokryje postupně. Upřesněte časové období, geografický rozsah, typ dat, o která stojíte (např. „*statistiky*“, „*porovnání*“, „*příčiny a dopady*“). Můžete zmínit, že chcete tabulky či grafy. Čím konkrétnější a strukturovanější dotaz, tím lepší strukturu bude mít odpověď.
- *Ověření a zodpovědné použití*: Ačkoli AI dokáže ohromně zrychlit získání informací, **není neomylná**. Vždy zůstává na uživateli, aby kriticky vyhodnotil získané výsledky. U reasoning modelů i DR reportů platí, že důležitá rozhodnutí by se neměla opírat slepě jen o AI výstup bez kontroly. Používejte citace – DR vám je dává jako záchytné body, tak je využijte a podívejte se, zda interpretace AI odpovídá zdroji. Při správném použití jsou však tyto nástroje nesmírně mocné a mohou vývojářům, výzkumníkům i dalším profesionálům **uvolnit ruce od rutinní práce** a umožnit jim soustředit se na vyšší úroveň problémů.

K dalšímu studiu: Pro zájemce o vizuální demonstraci doporučujeme video „**Introduction to Deep Research**“ od OpenAI (Mark Chen a kol., Tokyo, 2023), které ukazuje, jak DR agent postupuje a jaké výsledky generuje ([Introduction to Deep Research | IBL News .es](#)). Dále srovnávací článek *Perplexity vs OpenAI vs Google Deep Research* ([Perplexity Deep Research Takes on OpenAI & Gemini](#)) s praktickými ukázkami promptů a výstupů. Tyto zdroje poskytují hlubší náhled a konkrétní příklady, jež vám pomohou osvojit si práci s reasoning modely a DR nástroji v praxi.