Pokročilé techniky
Chain of Thought, Self-Reflection, ReAcT, RAG

Mgr. Igor Červený

Předmět: Využívání umělé inteligence (AI) jako nástroje pro práci VŠ pedagoga
Projekt: Podpora Rozvoje Učitelských Kompetencí (PRoUK)
Registrační číslo projektu: CZ.02.02.XX/00/23_019/0008385
Operační program: Jan Amos Komenský

Spolufinancováno Evropskou unií

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Kdy základní "řekni mi" nestačí?

- Základní prompty jsou skvělé pro rychlé informace, ale co když potřebujeme:
- Vyřešit komplexnější problém s více kroky?
- Zajistit vyšší přesnost a snížit riziko chyb?
- Aby AI sama zkontrolovala a vylepšila svůj výstup?
- Aby AI pracovala s aktuálními informacemi nebo specifickými dokumenty?

3. Chain of Thought (CoT)

- Model generuje myšlenkový postup krok za krokem (řetězec mezilehlých úvah) předtím, než sdělí konečnou odpověď. Místo okamžité odpovědi se model „zamyslí nahlas“.
- Proč "Řetězec myšlenek"?
- Protože model generuje sérii (řetězec) mezilehlých úvah, dílčích závěrů nebo výpočtů.
- Tyto jednotlivé "myšlenky" na sebe logicky navazují a společně tvoří cestu k řešení.

Analogie s lidským řešením problémů

- Když žák/student řeší složitou matematickou slovní úlohu nebo píše argumentační esej, obvykle:
- Analyzuje problém: Co se po mně chce? Co znám?
- Rozkládá ho na menší části: Jaké dílčí kroky musím udělat?
- Postupuje krok za krokem: Řeší jednotlivé části, odvozuje mezivýsledky.
- Skládá dílčí výsledky dohromady: Dochází k finálnímu řešení.
- (Ideálně) Prezentuje svůj postup: Ukazuje, jak k výsledku dospěl.

Analogie s lidským řešením problémů

- CoT vede AI k podobnému chování:
- Model se nesnaží "uhodnout" odpověď na komplexní otázku jedním skokem.
- Místo toho je veden k tomu, aby si problém strukturoval a postupně ho "odpracoval".

Analogie s lidským řešením problémů

- Zlepšení přesnosti u komplexních úloh:
- U úloh vyžadujících více kroků uvažování (matematika, logika, plánování, porozumění složitým textům) CoT výrazně zvyšuje šanci na správný výsledek.
- AI modely, i ty velmi pokročilé, mohou udělat chybu, pokud se snaží na složitou otázku odpovědět příliš rychle a přímo. CoT jim dává "prostor na přemýšlení".

Analogie s lidským řešením problémů

- Větší transparentnost a srozumitelnost:
- Vy (i vaši žáci) vidíte, jak AI k odpovědi dospěla.
- Umožňuje to lépe odhalit případné chyby v logice AI nebo nesprávné mezikroky.
- Didaktická hodnota:
- Ukázka "správného" myšlenkového postupu může sloužit jako vzor pro studenty.
- Můžete s žáky analyzovat jednotlivé kroky AI a diskutovat o nich.

Sebe-reflexe (Self-Reflection) – Co to je?

- Princip: Model nejprve vygeneruje odpověď a poté ji sám vyhodnotí a případně opraví. Iterativní smyčka.
- Liší se: Aktivace až po prvním pokusu, model si klade otázky ke kvalitě.
- Zlepšuje: Odhalení chyb, doplnění informací, přesnější výstup.
- Halucinace: Snížení rizika díky sebekontrole.
- Scénáře: AI jako "student" (píše, hodnotí, opravuje), AI jako "tutor" (vysvětluje, hodnotí vysvětlení, vylepšuje).

Jak to vypadá v praxi a klíčový rozdíl bez CoT

- Bez CoT (Zero-shot na složitou otázku):
- Uživatel: "Petr má 5 jablek, Eva mu dá 2x více jablek, než má Petr. Poté Petr sní 3 jablka. Kolik jablek Petrovi zbylo?"
- AI (může odpovědět rovnou, ale s rizikem chyby): "12" (Správně je (5 + 2*5) - 3 = 12, ale mohla by udělat chybu).

Jak to vypadá v praxi a klíčový rozdíl s CoT

- S CoT (explicitní pokyn nebo few-shot příklady s postupem):
- Uživatel: "Petr má 5 jablek, Eva mu dá 2x více jablek, než má Petr. Poté Petr sní 3 jablka. Kolik jablek Petrovi zbylo? Vysvětli svůj postup krok za krokem."
- AI:
- "Nejprve zjistíme, kolik jablek dala Eva Petrovi: Eva dala 2 * 5 = 10 jablek."
- "Poté sečteme původní počet Petrových jablek s jablky od Evy: Petr měl celkem 5 + 10 = 15 jablek."
- "Nakonec odečteme jablka, která Petr snědl: 15 - 3 = 12 jablek."
- "Petrovi tedy zbylo 12 jablek."

RAG: Jak to funguje a proč je to revoluční

- LLM vědí hodně, ale ne všechno (a ne vždy aktuálně)
- Znalostní "Cut-off Date":
- LLM jsou trénovány na obrovském množství dat, ale tato data mají časové omezení (poslední datum aktualizace).
- Nevědí o událostech, objevech nebo změnách, které nastaly po tomto datu.
- Halucinace a faktické chyby:
- Mohou si "vymýšlet" fakta nebo prezentovat zastaralé informace jako aktuální, protože nemají mechanismus pro ověření proti reálnému světu v reálném čase.
- Nedostatek přístupu ke specifickým/privátním datům:
- Standardní LLM nemají přístup k vašim konkrétním učebnicím, školním dokumentům, interním metodikám nebo nejnovějším odborným článkům, které nejsou součástí jejich trénovacích dat.

RAG: Chytré spojení LLM s externími znalostmi

- Retrieval Augmented Generation (Generování rozšířené o vyhledávání) je architektura, která kombinuje sílu velkých jazykových modelů (LLM) s externí databází znalostí.
- Představte si to jako kdyby LLM dostalo možnost "nahlédnout do knihovny" nebo do specifických dokumentů PŘEDTÍM, než začne generovat odpověď.
- Klíčový princip:
- Místo aby se LLM spoléhal jen na své interní (a potenciálně zastaralé) znalosti, RAG mu poskytne relevantní kontext z externího zdroje přímo v okamžiku dotazu.

Využití RAG ve vzdělávacím kontextu

- 1. Dotaz uživatele:
- Uživatel (např. pedagog, žák) položí otázku nebo zadá úkol.
- Příklad: "Jaké jsou nejnovější doporučení MŠMT pro využívání AI ve výuce?"
- 2. Vyhledávání (Retrieval):
- Systém RAG nejprve prohledá připojenou externí znalostní bázi (např. databázi dokumentů MŠMT, aktuální odborné články, vaše nahrané PDF soubory).
- Cílem je najít nejrelevantnější části textu (tzv. "chunks" nebo "context") související s dotazem.

Využití RAG ve vzdělávacím kontextu

- 3. Rozšíření (Augmentation):
- Nalezený relevantní kontext je automaticky přidán k původnímu dotazu uživatele.
- Vznikne tak "rozšířený" prompt pro LLM.
- Příklad rozšířeného promptu: "Na základě následujících informací z dokumentu MŠMT [zde je vložen relevantní text z dokumentu]... odpověz na otázku: Jaké jsou nejnovější doporučení MŠMT pro využívání AI ve výuce?"
- 4. Generování (Generation):
- Tento rozšířený prompt je odeslán do LLM.
- LLM nyní generuje odpověď, přičemž primárně vychází z poskytnutého relevantního kontextu, nikoliv jen ze svých obecných trénovacích dat.

Checklist pro skvělý prompt

- 🎯 Jasný cíl: Co přesně chcete, aby AI udělala?
- 👩‍🏫 Role (Persona): Kým má AI být? (učitel, expert, pětileté dítě...)
- 👥 Cílová skupina: Pro koho je výstup určen? (věk, úroveň znalostí)
- 📝 Formát výstupu: Odrážky, tabulka, báseň, počet vět/slov?
- 💡 Příklady (Few-shot): Ukažte AI, co chcete!
- 🚫 Omezení/Negace: Co AI NEMÁ dělat? (nepoužívej odborné termíny, vyhni se...)
- 🗣️ Tón: Formální, neformální, hravý, odborný?

CRAAP framework

- AI není neomylná: Proč je kritické myšlení klíčové?
- Jazykové modely (jako ty v Google AI Studiu) jsou úžasné nástroje, ale...
  - Mohou generovat **faktické chyby** ("halucinace").
  - Mohou obsahovat **zastaralé informace**.
  - Mohou odrážet **předpojatosti** (bias) z trénovacích dat.
  - Mohou znít **velmi přesvědčivě**, i když se mýlí.
- Pro pedagogy a žáky/studenty je zásadní:
  - **Nespolehnout** se slepě na první odpověď.
  - **Aktivně ověřovat** a kriticky posuzovat informace.

CRAAP Framework: 5 klíčových otázek

- Co je CRAAP? Mnemotechnická pomůcka pro kritéria hodnocení informací:
  - C – Currency (Aktuálnost)
  - R – Relevance (Vztah k tématu)
  - A – Authority (Důvěryhodnost zdroje/autora)
  - A – Accuracy (Přesnost)
  - P – Purpose (Účel)
- Proč je užitečný pro AI výstupy? Pomáhá nám systematicky identifikovat potenciální problémy a slabiny v generovaném obsahu.

Zdroje

- BRYNJOLFSSON, Erik; MCAFEE, Andrew a DRLÍK, Filip. *Druhý věk strojů: práce, pokrok a prosperita v éře špičkových technologií*. V Brně: Jan Melvil Publishing, 2015. ISBN 978-80-87270-71-4.
- COECKELBERGH, Mark a FICOVÁ, Sylva. *Etika umělé inteligence*. Praha: Filosofický ústav AV ČR ve svém nakladatelství Filosofia, 2023. ISBN 978-80-7007-746-7.
- ÖZKIZILTAN, Didem. Melanie Mitchell: Artificial intelligence—a guide for thinking humans: Picador, New York, 2019, ISBN: 978-1-250-75804-0.
- TAULLI, Tom. *Artificial intelligence basics: a non-technical introduction*. Monrovia, California: Apress, 2019. ISBN 1-4842-5028-1.
- Wooldridge, Michael J. *A brief history of artificial intelligence: what it is, where we are, and where we are going*. First U.S. edition. New York: Flatiron Books, 2021. vii. ISBN 978-1-250-77074-5.

Otázky / Debata

Děkuji za Vaši pozornost

Mgr. Igor Červený

Spolufinancováno Evropskou unií

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY