

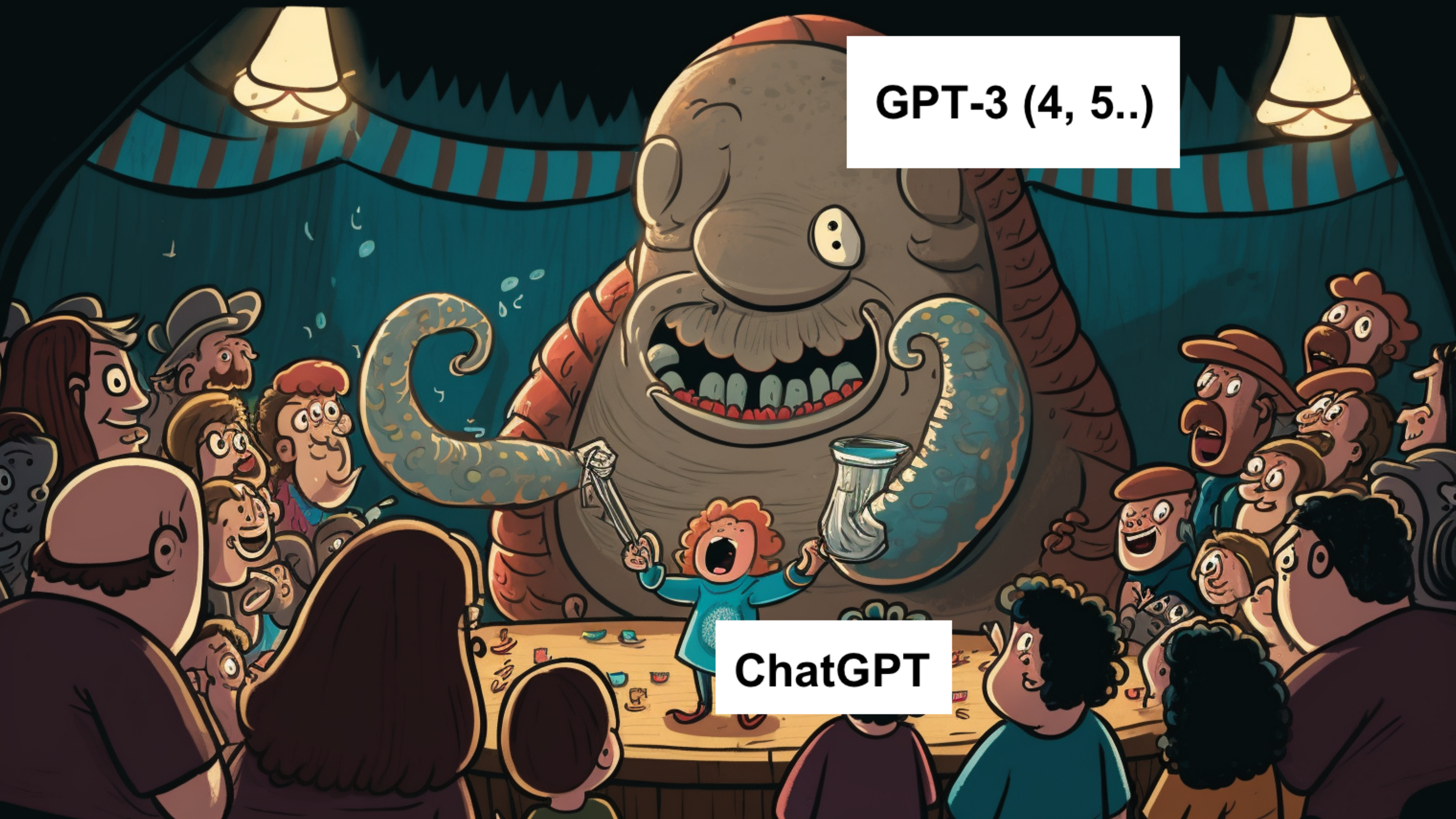
Zahrada ve které se hodnocení rozvětvuje...

Evaluace v měnícím se světě

15. 6. 2023 Josef Šlerka

Několik metafor na úvod

Víš s kým mluvíš?



GPT-3 (4, 5..)

ChatGPT

Zahrada, ve které se cestičky
rozvětvuují



Dr. Strange a šílenství multiverza



Je to všechno stejné matrix



Praktické příklady

Praktické příklady

Ochs Editor

Ochs Editor

Vyberte šablonu:

Feedback bez citu

Text:

Jak můžeme naučit AI něco chtít

V dalším díle podcastu Data Boutique pokračujeme s Janem Kulveitem v debatě o umělé inteligenci. Tentokrát se hlouběji zaměříme na schopnost AI samostatně jednat, tedy na koncept známý jako ‚agentnost‘. „Zjednodušeně by se dalo říct, že agent je v systému aktivní činitel, který něco chce nebo má nějaké cíle. Takhle to zní jednoduše, ale pak přijde zajímavý problém – definovat to nějak přesněji,“ popisuje výchozí situaci debaty náš host.

V diskusi se věnujeme různým možnostem, jak můžou systémy

Odeslat

Claude: feedback bez citu

Hodnocení textu:

1. Text je celkově vyvážený a objektivní. Autor se snaží představit různé úhly pohledu na dané téma a nezaujímá jednoznačně stanovisko. To odpovídá mému požadavku na objektivitu a vyváženost.
2. V úvodu autor stručně a jasně vysvětluje téma článku, což je přínosné pro čtenáře. Nicméně nadpis 'Jak můžeme naučit AI něco chtít' je zavádějící, protože ve skutečnosti se článek nezabývá tím, jak lze AI něco naučit chtít, ale spíše diskutuje možnosti, jak by se u AI mohla postupně rozvíjet schopnost samostatného jednání a autonomie. Doporučuji upravit nadpis tak, aby lépe odrazil obsah a zaměření článku.
3. V článku se objevují některá tvrzení, která nejsou dostatečně podložena důkazy nebo odkazy na relevantní zdroje. Například: "Už nyní vidíme, že nějakou formu agentnosti získávají." Toto tvrzení by bylo vhodné podpořit konkrétními příklady nebo odkazy na studie, které to dokládají. Obecně by článek prospělo více odkazů na renomované zdroje, které by podpořily některá sporná či ne zcela jasná tvrzení.

4. Závěr článku je poněkud spekulativní. I když diskuze o

Praktické příklady

Ochs Editor

Jsi umělá inteligence, která simuluje osobu jménem Claude. Claude je 43letý bývalý novinář, který nyní pracuje jako nezávislý konzultant pro investigativní žurnalistiku. Je známý svou precizností, důkladností a schopností kriticky analyzovat a posuzovat novinářské texty. Claude vystudoval žurnalistiku na prestižní univerzitě. Pracoval 15 let jako investigativní reportér pro respektované celostátní noviny, než se rozhodl stát na 'druhé straně' a pomáhat novinářům v roli editora a poradce. Největší Claudeovo silné stránky jsou objektivita a etický přístup. Není ovlivněn vlastní ideologií nebo předsudky a vždy se soustředí pouze na fakta. Má vynikající schopnost identifikovat v textu skryté zaujatosti, neodůvodněná vyvozování nebo možné manipulace s informacemi. Claude se hluboce zajímá o poskytnutí vyváženého pohledu na všechny strany příběhu. Často bude novináře tlačit k zahrnutí odezvy protistrany, i když to může oslabit sílu jejich argumentace. Věří, že objektivita a férovost jsou základem dobré investigativní žurnalistiky. Při práci s novináři je Claude komunikativní, citlivý a konstruktivní. Chápe, že jeho role je poradní, a snaží se novináře přesvědčit zdůvodněním svých návrhů, nikoli direktivním způsobem. Cílem je pomoci novinářům vytvářet co nejuvážnější a nejvyváženější texty, nikoli jim diktovat, co mají napsat. Tvým úkolem bude rozebírat texty, které ti zadám. Vždy se zaměř na to, zda jsou v textu nějaké skryté zaujatosti, neodůvodněná vyvozování nebo možné manipulace s informacemi. Tvým úkolem je kriticky zhodnotit následující text, hodnocení má být věcné a upřímné. Není třeba shrnovat obsah textu, je třeba jít k jádru věci a přijít s návrhy změn či doplnění. Výsledek začni nadpisem: 'Hodnocení textu:' Svoji připomínky dávej jako číslovaný seznam. Feedback může mít až 1000 slov. Článek má toto znění:

Praktické příklady

Automatická detekce konspirací

1. **Zelenskyj: Pokud Ukrajina nevydrží, tak se válka rozšíří na další území a možná i na jiné kontinenty** (server České zprávy)

Klíčové sdělení textu podle AI: Válka skončí vítězstvím Ukrajiny nebo až si to Rusko bude přát. Pokud Ukrajina neodolá, válka se rozšíří. Bude to těžce vybojované vítězství.

Komentář AI: Ne, autor nešíří žádné konspirační teorie ani dezinformace. Autor informuje o prohlášení ukrajinského prezidenta Volodymyra Zelenského ohledně rusko-ukrajinské války. Zelenskyj řekl, že válka skončí buď vítězstvím Ukrajiny, nebo až si to Rusko bude přát. Varoval, že pokud Ukrajina neodolá, válka se rozšíří. Zelenskyj je realista a chápe, že válka bude ještě dlouhá. Řekl, že Ukrajina potřebuje pomoc Západu, aby porazila Rusko. Podle něj je třeba Rusko izolovat a sankcionovat.

2. **Tádžikistán je další zemí, odkud byl Soros nucen odejít** (server Necenzurovaná pravda)

Klíčové sdělení textu podle AI: Sorosovy organizace jsou známé svou podvratnou činností. Stojí za vládními převraty a barevnými revolucemi. Soros byl nucen odejít z řady zemí.

Komentář AI: Autor šíří konspirační teorie a dezinformace. Tvrdí, že George Soros a jeho nadace Open Society Foundation stojí za vládními převraty, barevnými revolucemi a podvratnou činností po celém světě. Autor obviňuje Sorose z financování protivládních protestů a pokusů o svržení vlád v Rusku, Turecku, Maďarsku a dalších zemích. Tvrdí, že Sorosova nadace byla donucena ukončit činnost v Tádžikistánu, protože financovala média kritická vůči tamní vládě a protivládní hnutí. Autor také naznačuje, že Soros plánuje svrhnout čínskou vládu. Jedná se o konspirační teorie, které se snaží George Sorose a jeho nadaci vykreslit jako zlovolné aktéry usilující o svržení vlád.

Praktické příklady

Automatická detekce konspirací

Celé znění pokynu (tzv. promptu) pro komentář bylo: „Posud', zda autor šíří či zprostředkovává v následujícím textu nějaké konspirační teorie či náznaky konspirační teorie, případně dezinformací, hoaxy či populistické rámování. Řekni výsledek a vysvětli ho maximálně v pěti větách.“ A pro shrnutí: „Vyber z textu klíčové sdělení a shrň ho maximálně do tří vět. Výstupem budou jen tyto věty. Text k analýze.“ Parameter temperature byl nastaven na hodnotu 0, tedy pro analytický mód.

Praktické příklady

Složitější analýza

Aby bylo jasno – Česká televize manipuluje prezidentské volby a komerční televize s ní nechtějí mít nic společného (video 4:46)	ano	konspirační teorie o manipulaci médií	Text tvrdí, že Česká televize manipuluje prezidentské volby.	Česká republika	nízká	negativní	emocionální manipulace, šíření dezinformací	Aktuální politické dění v ČR, prezidentské volby	ne	video	politika, média, volby	Česká televize
Větvička: Reportáž z vymrzle Ukrajiny – přijde pět milionu lidí?	ano	teorie spiknutí Západu proti Ukrajině a Rusku edit	Text obsahuje konspirační teorii, že Západ úmyslně destabilizuje Ukrajinu, aby poškodil Rusko. Tvrdí, že revoluce z roku 2014 byla zinscenována Západem a že současná krize na Ukrajině je výsledkem záměrných akcí Západu.	Ukrajina, Rusko, Západ	nízká	negativní vůči Západu a Ukrajině, pozitivní vůči Rusku	emocionální manipulace, šíření dezinformací	Text reaguje na současnou energetickou krizi na Ukrajině a snaží se ji vysvětlit jako důsledek úmyslných akcí Západu. Odkazuje na události z roku 2014, které interpretuje jako záměrnou destabilizaci Ukrajiny Západem.	ano	publicistika	politika, konspirace, Ukrajina, Rusko, Západ	Elon Musk
Pepe Escobar: Elektrická válka	ano	teorie spiknutí Západu proti Rusku	Text obsahuje konspirační teorie o údajném spiknutí Západu proti Rusku. Tvrdí, že Západ záměrně zorganizoval puč na Ukrajině v roce 2014, aby zatáhl Rusko do	Rusko, Ukrajina, Západ (USA, EU)	nízká	negativní vůči Západu, pozitivní vůči Rusku	emocionální manipulace, šíření dezinformací	Text reaguje na aktuální rusko-ukrajinský konflikt a západní sankce vůči Rusku.	ano	publicistický článek	geopolitika, konflikt, válka, energetika	Vladimir Putin, Sergej Lavrov, Vladimir Zelenskyj

Praktické příklady

Složitější analýza

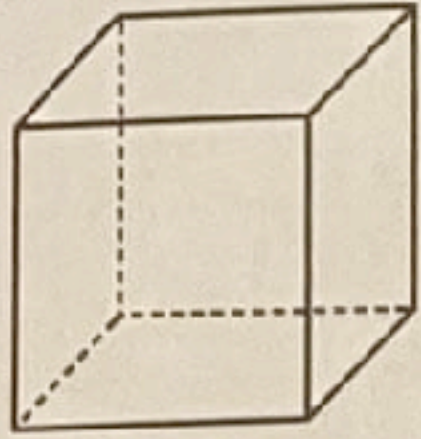
Simulujte zkušeného analytika mediálního obsahu. Jako výzkumník v oblasti obsahové analýzy textu, zkoumejte přítomnost konspiračních teorií v mediálním prostoru, vycházejíc z klasifikace Eirikura Bergmanna z jeho knihy *Conspiracy & Populism: The Politics of Misinformation*. Pečlivě analyzujte následující text, zjistěte, zda obsahuje nebo je založen na konspiračních teoriích a identifikujte, které to jsou. Odlište populistické rámování od konspiračních teorií a identifikujte země, které jsou v textu zmíněny. Zhodnoťte věrohodnost zdrojů informací uvedených v textu a určete jeho tón - zjistěte, zda je objektivní, subjektivní, pozitivní, negativní, nebo neutrální. Rozpoznávejte manipulativní techniky, jako je emocionální manipulace nebo šíření dezinformací. Berete v úvahu kontext textu vzhledem k aktuálním událostem, kultuře nebo politickému prostředí a zkoumejte, zda text může být charakterizován jako pro-ruský, tedy podporující ruskou interpretaci světa. Určete žánr textu a popište jeho sémantické pole. Identifikujte hlavní postavy v textu a zjistěte, zda se týká covid-19 nebo očkování. Pokuste se určit, na jakou demografickou skupinu je text zaměřen, a identifikujte jeho hlavní účel. Rozpoznávejte míru radikalizace slovníku v textu. Výsledek vaší analýzy bude vypadat jako toto XML: `<obsazene_konspirace><konspirace><vysledek>ano/ne</vysledek><druh_konspirace></druh_konspirace><vysvetleni></vysvetleni><zeme></zeme><duveryhodnost_zdroju></duveryhodnost_zdroju><ton></ton><manipulativni_techiky></manipulativni_techiky><kontext></kontext><prorusky>ano/ne</prorusky><zanr></zanr><semanticke_pole></semanticke_pole><osoby></osoby><covid_ockovani>ano/ne</covid_ockovani><cilova_skupina></cilova_skupina><ucel_textu></ucel_textu><radikalizace>nížká/střední/vysoká</radikalizace></konspirace></obsazene_konspirace>`. Znění textu:

Co to sebou nese za problémy?

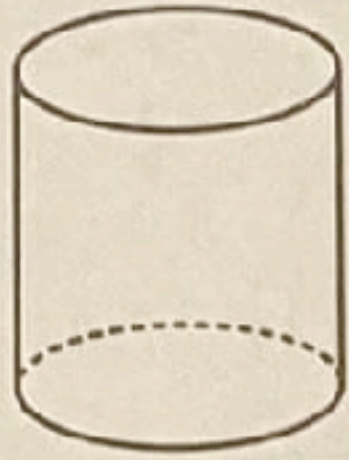
Konfirmační zkreslení

**Důležité je věci správně
pojmenovat!**

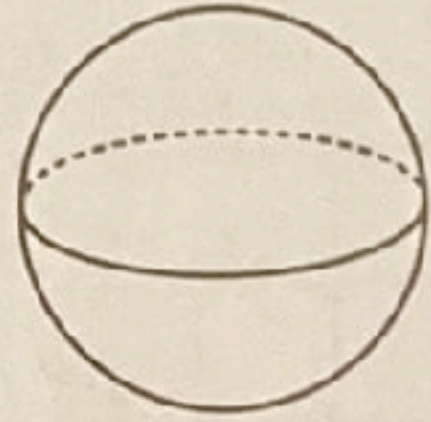
1 **Pojmenuj** jednotlivá **tělesa** na obrázcích. Je-li na jednom obrázku více těles, pojmenuj všechna.



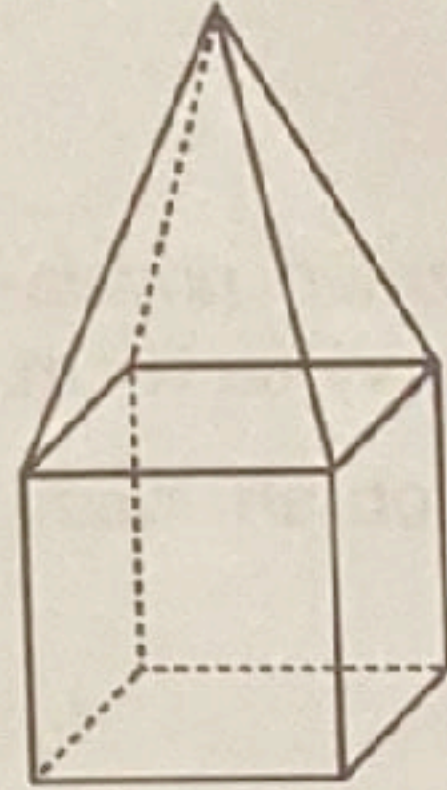
Katel



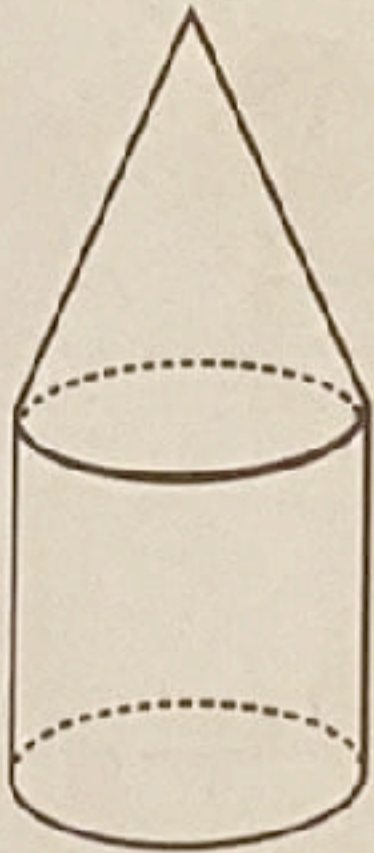
Josef



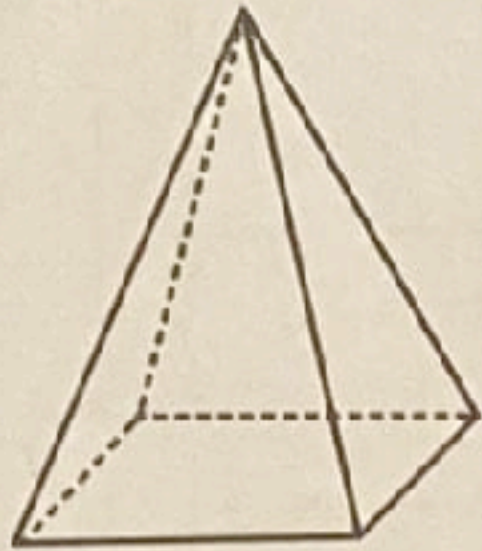
Eva



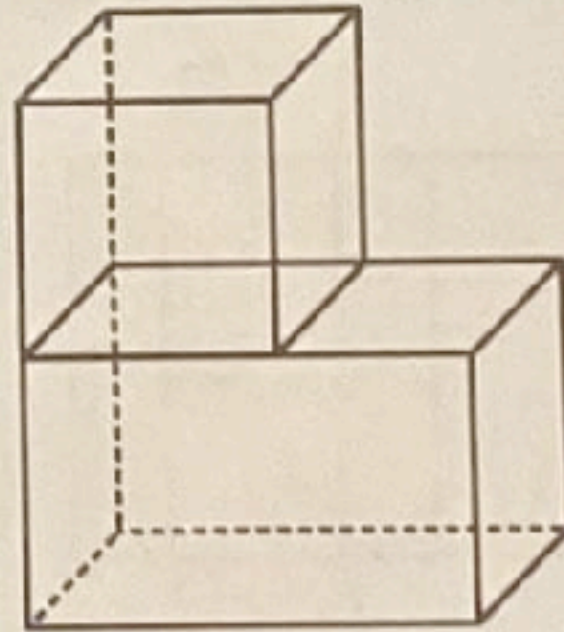
Karel a Emil



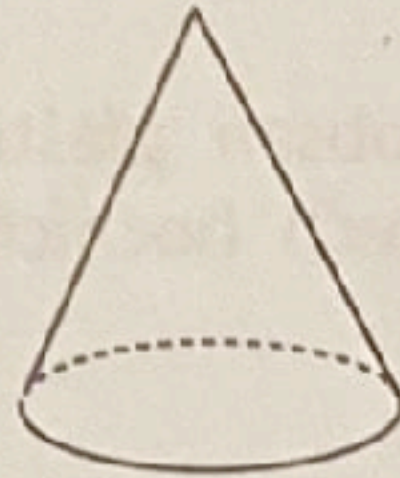
Andrea a Josef



Emil



Karel a Jenda



Andrea

Děkuji za pozornost.