

Generování reportu z odehraného zápasu

Generaci krátkého článku můžeme dělat několika způsoby.

Generace textu pomocí LLM

Mohly bychom využít některý z před-natréovaných generativních modelů (např. GPT) s promptem odpovídající zadání této úlohy s vloženými informacemi o zápasu, formátovány tak aby to náš model správně pochopil (např. „Skóre bylo 2-0“, „První gól byl skórován týmem A v prvním poločase“). Poté u výsledného textu ověříme výskyt požadovaných informací, například tak, že zkontrolujeme, zda se v textu objevují jména hráčů, datum zápasu, čísla odpovídající výsledku atd. Pokud by klíčová slova nebyla nalezena zkusily bychom článek vygenerovat znovu s trochu jiným (předpřipraveným) promptem. Výstup bychom pro jistotu ještě měli nechat zkontrolovat člověkem, abychom si byly jisti, že nám model nevygeneroval nesmysl.

Správnost modelu bychom mohli zvýšit finetuningem, kde bychom natrénovali model na lidsky napsaných shrnutích. Toto by ale bylo efektivní pouze pokud bychom už hodně takových shrnutích měli, jinak bychom museli najmout člověka, aby je pro trénink napsal manuálně, ale v tom případě by bylo efektivnější rovnou na několik let tohoto člověka najmout, aby shrnutí psal přímo.

Efektivita tohoto přístupu závisí na kvalitě použitého modelu a jeho finetuningu. Za nejkvalitnější modely se platí, ale na generaci takto relativně krátkých článků může stačit i

Generace textu pomocí vzorů

Abychom vygenerovaly požadovaný text, tak by stačilo vytvořit několik vzorových vět, do kterých bychom poté požadovanou informaci vložily.

Vzorová věta by vypadala například takto: „Tým A porazil team B s výsledkem X:Y.“

Při generaci by znaky A, B, X a Y byly nahrazeny informacemi odpovídající našemu zápasu.

Museli bychom napsat několik (10-20) vět pro každou informaci, aby nebyly články příliš repetitivní.

Tento přístup má tu výhodu, že máme vždy garantováno, že se v textu objeví požadované informace, ale za cenu toho, že celkový článek nebude mít globální strukturu, ve smyslu toho, že bude obtížné vytvořit příběh o průběhu hry, např. tuto větu by bylo obtížné vytvořit: „V **druhém poločase** skóroval tým B druhý gól, **čímž obrátil průběh hry**“. Samozřejmě bychom vzor této věty mohli vytvořit, za přidání podmínek že se skórovalo v druhém poločase a skóre bylo X:0. Ale pokud bychom chtěli hodně takovýchto kontextově závislých vět, tak by jejich psaní a vytváření podmínek mohlo zabrat podstatné množství práce, a i pak by možná mohl text být špatně čitelný, pokud by se věty za sebe nevhodně naskládaly.

Hybridní přístup

Mohli bychom zkombinovat tyto dva přístupy, abych zkombinovali jejich výhody. Jako základ bychom použily Vzorový přístup, ale věty, které by používal by byly přímo generovány LLM modelem, místo toho, aby byly předepsány. LLM model by měl za vstup již vygenerovaný text a mohl by tedy v nové větě psát o předcházejících událostech, zároveň, protože chceme vygenerovat pouze jednu větu s jedním kouskem informace, je mnohem pravděpodobnější, že náš model požadovanou informaci skutečně napíše. Hybridní přístup bude pravděpodobně nejlepší z těchto tří přístupů.