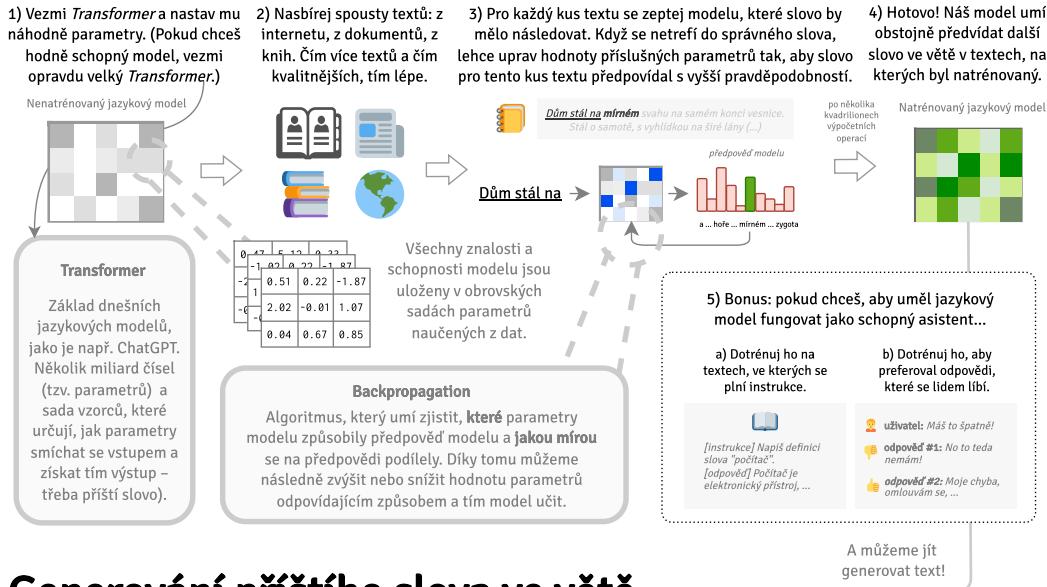
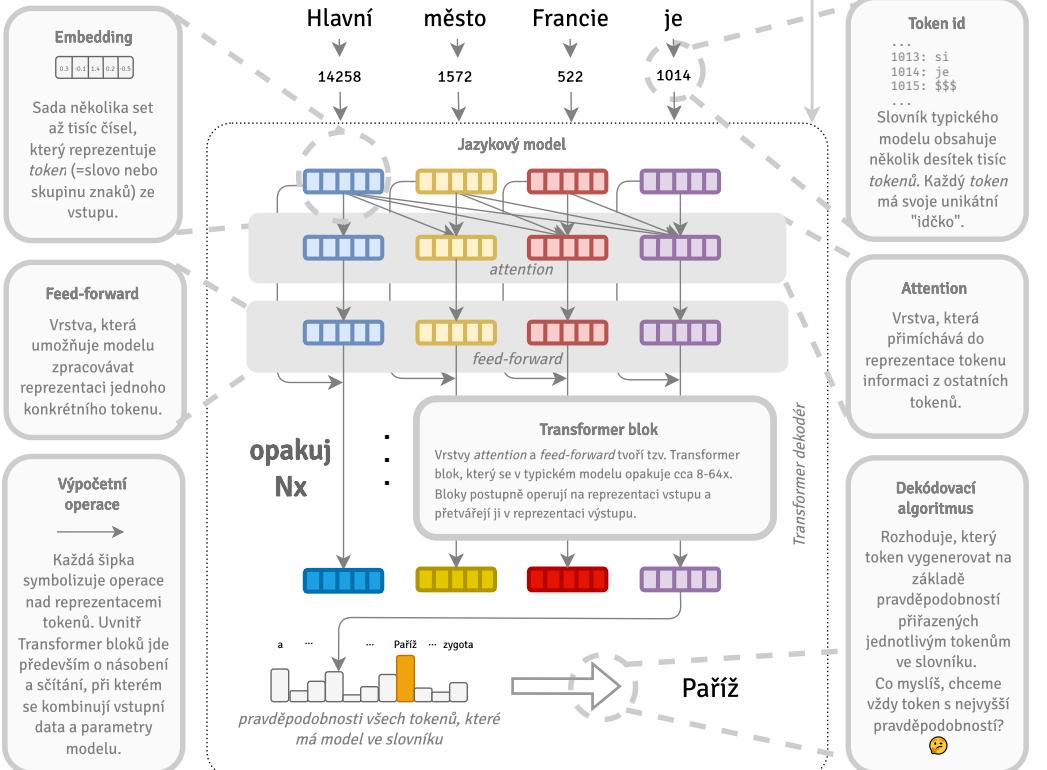


Recept na model, který pohání ChatGPT



Generování příštího slova ve větě



NOVĚ: Toto vše najdete animované na <https://animatedllm.github.io/>!

Generování textu: otázky a odpovědi

Umí si jazykové modely i hledat v internetu?



Krátká odpověď: Neumí, ale někdo to může dělat za ně.

Dlouhá odpověď: Jazykový model jako takový hledat na internetu neumí. Všechno, co umí, se musel naučit během trénování.

V dnešní době ale mají komerční jazykové modely typicky externí vyhledávací modul. Tento modul hledá informace související s dotazem a předkládá je jazykovému modelu na vstupu společně se vstupem od uživatele. Model pak s nimi pracuje stejně, jako se zbytkem textu. Kvalita těchto výsledků může ovlivnit vygenerovaný text jak pozitivně, tak negativně.

Má model seznam všech slov ve všech jazycech?



Krátká odpověď: Ne – ale jejich částí ano!

Dlouhá odpověď: Všechna slova bylo opravdu až příliš (už jen všechny vyskloňované tvary slov v češtině!). Proto používáme jako tokeny tzv. "subwordy": slova a části slov, se kterými můžeme libovolná slova poskládat. Frekventovanější slova máme ve slovníku přímo, ta méně častá složíme z více částí.

Rozsekáme text na subwordy a pak ho zase poskládat je úkolem specializovaného algoritmu, tzv. tokenizéra, který pracuje nezávisle na jazykovém modelu.

Jak energeticky náročné je natřenovat model? A kolik energie spotřebuje jeden dotaz?

Krátká odpověď: Trénovat je náročné, generovat ne tolik.

Dlouhá odpověď: Záleží na velikosti modelu, ale odhaduje se, že natřenovat model s 175 mld. parametry stojí 1.2 GWh energie, což odpovídá roční spotřebě 120 amerických domácností. Modely se ale naštěstí netrenují tak často.

Generovat na natřenovaném modelu text už je mnohem efektivnější: průměrná odpověď ChatGPT spotřebuje cca 0,3 Wh, což odpovídá např. rychlovárně konvici zapnuté na 1 sekundu.

[1] <https://blog.camniran.com/the-point-of-singularity>
[2] <https://idioti.org/how-much-energy-do-lms-consume-unveiling-the-power-behind-ai/>
[3] <https://andymasley.substack.com/p/a-cheat-sheet-for-conversations-about>

Co dělá velké jazykové modely tak inteligentní?



Krátká odpověď: To je mi ale těžká otázka...!

Dlouhá odpověď: Modely do nějaké míry umějí zopakovat to, co viděly v trénovacích datech. Chytrá odpověď proto mohla jen "ležet na internetu".

Modely ale umí částečně i generalizovat: kombinovat naučené vzory v originální odpovědi. Důvodem může být to, že s velkým množstvím dat může být generalizovat jednodušší, než se učit vše nazepamět.

A tady už přichází další otázky: Co dělá inteligentní člověka? A co je to vlastně ta "intelligence"?

Kde se můžu naučit o jazykových modelech více?

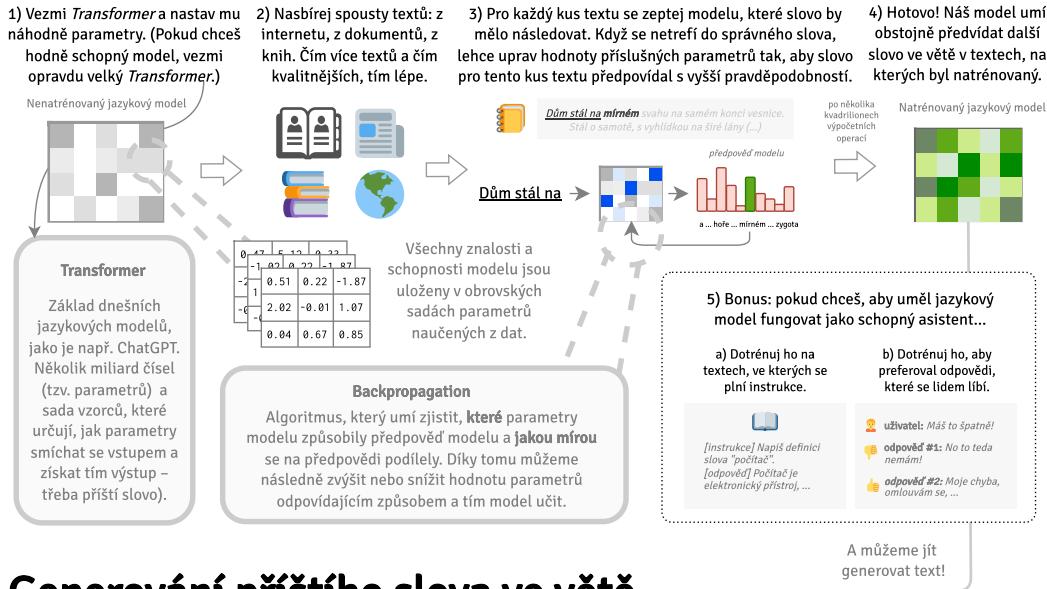


Krátká odpověď: U nás na ÚFALu!

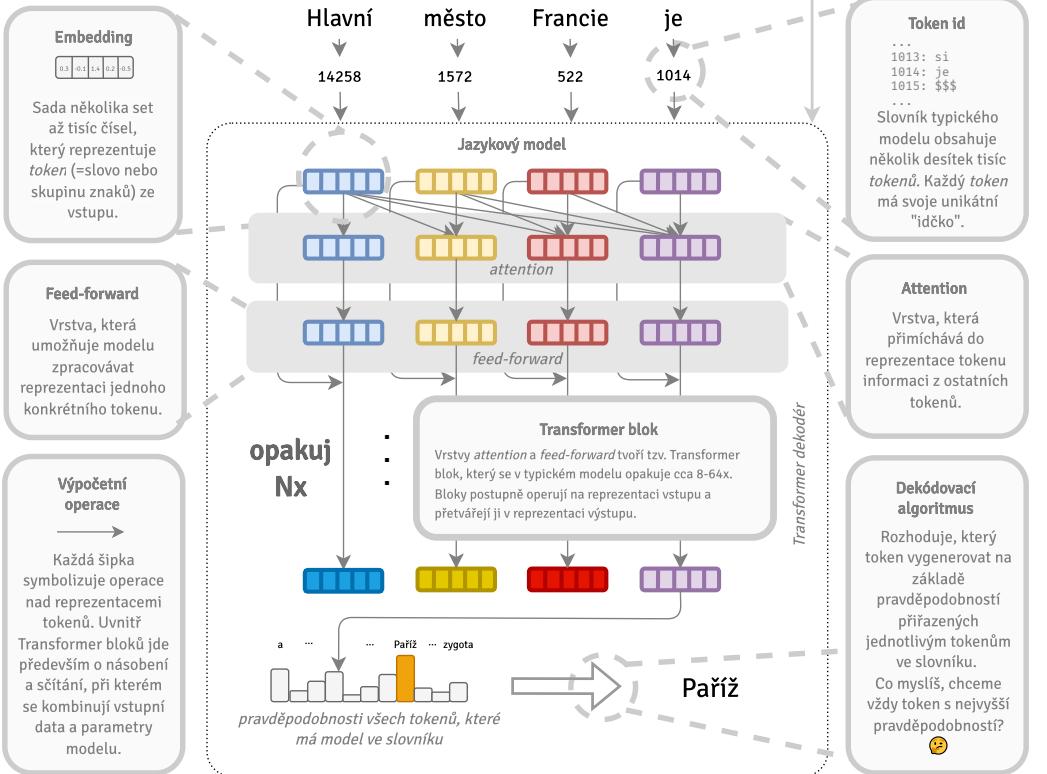
Dlouhá odpověď: Podobnými tématy se zabýváme na Ústavu Formální a Aplikované Lingvistiky (ÚFAL), což je součást Matematicko-fyzikální fakulty Univerzity Karlové.

Na naší katedře propoujeme znalosti o jazyce se znalostmi ze strojového učení. Učíme o tom bakalářské i magisterské předměty, jako třeba Deep Learning nebo Large Language Models.

Recept na model, který pohání ChatGPT



Generování příštího slova ve větě



NOVĚ: Toto vše najdete animované na <https://animatedllm.github.io/>!

Generování textu: otázky a odpovědi

Umí si jazykové modely i hledat v internetu?



Krátká odpověď: Neumí, ale někdo to může dělat za ně.

Dlouhá odpověď: Jazykový model jako takový hledat na internetu neumí. Všechno, co umí, se musel naučit během trénování.

V dnešní době ale mají komerční jazykové modely typicky externí vyhledávací modul. Tento modul hledá informace související s dotazem a předkládá je jazykovému modelu na vstupu společně se vstupem od uživatele. Model pak s nimi pracuje stejně, jako se zbytkem textu. Kvalita těchto výsledků může ovlivnit vygenerovaný text jak pozitivně, tak negativně.

Má model seznam všech slov ve všech jazycech?



Krátká odpověď: Ne – ale jejich částí ano!

Dlouhá odpověď: Všechna slova bylo opravdu až příliš (už jen všechny vyskloňované tvary slov v češtině!). Proto používáme jako tokeny tzv. "subwordy": slova a části slov, ze kterých můžeme libovolná slova poskládat. Frekventovanější slova máme ve slovníku přímo, ta méně častá složíme z více částí.

Rozsekáme text na subwordy a pak ho zase poskládat je úkolem specializovaného algoritmu, tzv. tokenizéra, který pracuje nezávisle na jazykovém modelu.

Jak energeticky náročné je natřenovat model? A kolik energie spotřebuje jeden dotaz?

Krátká odpověď: Trénovat je náročné, generovat ne tak.

Dlouhá odpověď: Záleží na velikosti modelu, ale odhaduje se, že natřenovat model s 175 mld. parametry stojí 1.2 GWh energie, což odpovídá roční spotřebě 120 amerických domácností. Modely se ale naštěstí netřenují tak často.

Generovat na natřenovaném modelu text už je mnohem efektivnější: průměrná odpověď ChatGPT spotřebuje cca 0,3 Wh, což odpovídá např. rychlovárně konvici zapnuté na 1 sekundu.

[1] <https://blog.camniran.com/the-point-of-singularity>
[2] <https://idioti.org/how-much-energy-do-lms-consume-unveiling-the-power-behind-ai/>
[3] <https://andymasley.substack.com/p/a-cheat-sheet-for-conversations-about>

Co dělá velké jazykové modely tak inteligentní?



Krátká odpověď: To je mi ale těžká otázka...!

Dlouhá odpověď: Modely do nějaké míry umějí zopakovat to, co viděly v trénovacích datech. Chytrá odpověď proto mohla jen "ležet na internetu".

Modely ale umí částečně i generalizovat: kombinovat naučené vzory v originální odpovědi. Důvodem může být to, že s velkým množstvím dat může být generalizovat jednodušší, než se učit vše nazepamět.

A tady už přichází další otázky: Co dělá inteligentní člověka? A co je to vlastně ta "intelligence"?

Kde se můžu naučit o jazykových modelech více?



Krátká odpověď: U nás na ÚFALu!

Dlouhá odpověď: Podobnými tématy se zabýváme na Ústavu Formální a Aplikované Lingvistiky (ÚFAL), což je součást Matematicko-fyzikální fakulty Univerzity Karlové.

Na naší katedře propoujdejeme znalosti o jazyce se znalostmi ze strojového učení. Učíme o tom bakalářské i magisterské předměty, jako třeba Deep Learning nebo Large Language Models.