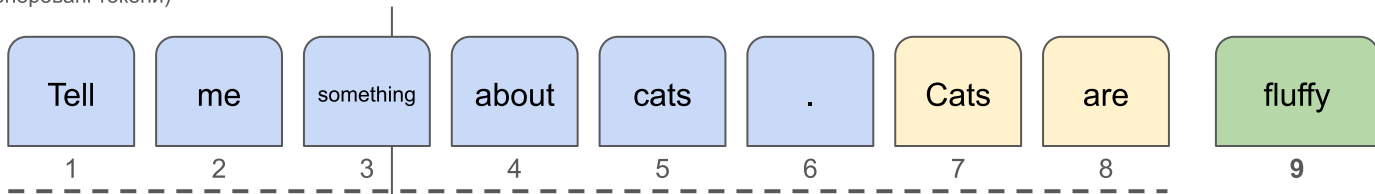


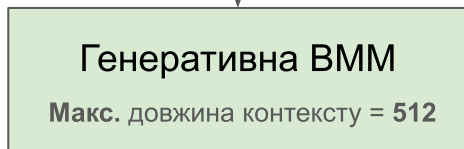
Вхідні дані:  
(запит + згенеровані токени)

Tell me something about cats.

Попередньо зген. токени



Поточна довжина контексту = 8



Відповідь моделі:  
(генерація по токеноу)



Оновлення відповіді  
(збільшує довжину контексту)