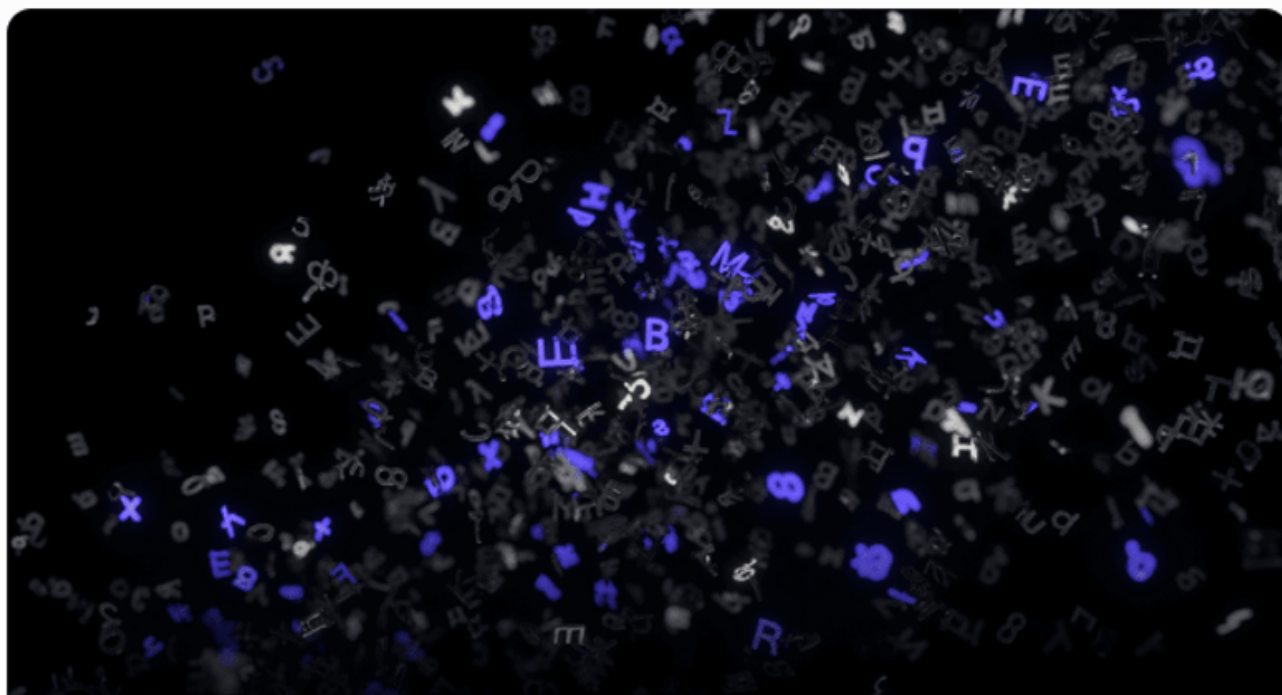


УКРАЇНСЬКА LLM НА ОСНОВІ GEMMA ВІД GOOGLE

Українську LLM натренують на моделі Gemma від Google



Міністерство цифрової трансформації України та Київстар повідомляють, що ними обрано велику мовну модель, на якій тренуватимуть національну українську LLM.

Нею стала загальнодоступна open-source-модель Gemma 3 від Google. Її планується адаптувати до української мови, удосконалити український токенайзер (це покращить роботу моделі з українською мовою, зменшить помилки при створенні україномовних текстів та оптимізує обчислювальні витрати при використанні моделі); донавчити модель на унікальних україномовних текстах, які зараз збирають експерти. Ключовими перевагами у виборі моделі стали наступні: оптимальний баланс продуктивності та ресурсів (Gemma забезпечує високу якість при оптимальних інфраструктурних вимогах, це одна з найкращих серед відкритих моделей з погляду співвідношення розміру та якості); багатомовна підтримка (модель наразі має українську мову у своєму діапазоні та легко адаптується через донавчання); мультимодальність (модель може сприймати й аналізувати не лише текст, а й зображення); розширений токенайзер (обсяг токенів забезпечує точне та ефективне опрацювання текстів та донавчання); довге контекстне вікно (128 тис. токенів); наявність кількох розмірностей (що дозволяє гнучко обирати

розмір моделі під конкретну сферу застосування); успішні приклади використання для створення українськомовних LLM (Lara LLM та MatayLM).

Детальніше: <https://qrcode.net/qr/2Orbs>, <https://qrcode.net/qr/jOKXY>

Фото: скріншот

#НРАТ_Усі_новини #НРАТ_ШтучнийІнтелект #НРАТ_Науковцям_новини
#НРАТ_Освітянам_новини #НРАТ_Інноваторам_новини #НРАТ_Бібліотекарям

2025-12-05

Інформація з офіційного вебпорталу Національного репозитарію академічних текстів