

ПОЯСНЮВАЛЬНА ЗАПИСКА
до проекту постанови Кабінету Міністрів України
“Деякі питання розвитку українських великих мовних моделей”

1. Мета

Проект постанови Кабінету Міністрів України “Деякі питання розвитку українських великих мовних моделей” (далі – проект акта) розроблено з метою розвитку українських великих мовних моделей.

2. Обґрунтування необхідності прийняття акта

На сьогодні використання великих мовних моделей в Україні переважно здійснюється із застосуванням іноземних технологічних платформ, що не забезпечує повною мірою врахування особливостей української мови, культури та історичного контексту.

У зв’язку з цим виникла потреба у створенні українського корпусу даних (далі – корпус даних) для навчання, донавчання, тренування та тестування великих мовних моделей та українських великих мовних моделей.

Створення корпусу даних передбачає формування систематизованої та структурованої сукупності наборів даних, що, зокрема, містять тексти українською мовою та інформацію, яка відображає мовні, культурні та історичні особливості України.

Для створення та наповнення корпусу даних необхідним є залучення наборів даних, що перебувають у розпорядженні органів державної влади, інших державних органів, державних наукових організацій та установ, архівних установ, бібліотек, закладів освіти, а також інших юридичних та фізичних осіб, фізичних осіб - підприємців.

Водночас потребує нормативного визначення механізм створення, наповнення та підтримання в актуальному стані корпусу даних, порядок надання наборів даних постачальниками даних, а також умови надання та припинення доступу до корпусу даних користувачам.

З огляду на зазначене та з метою розвитку українських великих мовних моделей пропонується визначити механізм створення, наповнення, підтримання в актуальному стані українського корпусу даних та надання доступу до нього.

3. Основні положення проекту акта

Проектом акта пропонується:

затвердити Положення про український корпус даних;

визначити механізм створення, наповнення, підтримання в актуальному стані українського корпусу даних та надання доступу до нього;

визначити порядок надання наборів даних для створення та наповнення українського корпусу даних;

визначити повноваження державного підприємства “ДІА” (укладача) щодо створення, наповнення, підтримання в актуальному стані українського корпусу даних, надання та припинення доступу до нього;

встановити вимоги до отримувачів доступу, умови надання доступу до українського корпусу даних та його використання користувачами;

визначити умови використання українського корпусу даних для навчання, донавчання, тренування та тестування великих мовних моделей та українських великих мовних моделей.

4. Правові аспекти

Проект акта розроблено відповідно до Положення про Міністерство цифрової трансформації України, затвердженого постановою Кабінету Міністрів України від 18 вересня 2019 р. № 856.

5. Фінансово-економічне обґрунтування

Реалізація проекту акта не потребує додаткових бюджетних коштів та не матиме прямого чи опосередкованого впливу на надходження та витрати державного та/або місцевого бюджетів.

6. Позиція заінтересованих сторін

Проект акта потребує проведення громадського обговорення відповідно до вимог Порядку проведення консультацій з громадськістю з питань формування та реалізації державної політики, затвердженого постановою Кабінету Міністрів України від 03 листопада 2010 р. № 996 “Про забезпечення участі громадськості у формуванні та реалізації державної політики”.

Проект акта не стосується сфери наукової та науково-технічної діяльності.

7. Оцінка відповідності

У проекті акта відсутні положення, що стосуються прав та свобод, гарантованих Конвенцією про захист прав людини і основоположних свобод, які впливають на забезпечення рівних прав та можливостей жінок і чоловіків, містять ризики вчинення корупційних правопорушень та правопорушень, пов’язаних з корупцією, створюють підстави для дискримінації, стосуються інших ризиків та обмежень, які можуть вплинути під час реалізації проекту акта.

Громадська антикорупційна, громадська антидискримінаційна та громадська гендерно-правова експертизи не проводились.

Проект акта містить положення, що стосуються зобов’язань України у сфері європейської інтеграції.

8. Прогноз результатів

Прийняття проекту акта забезпечить розвиток українськомовних технологій штучного інтелекту.

Реалізація акта матиме вплив на інтереси заінтересованих сторін:

Заінтересована сторона	Вплив реалізації акта на заінтересовану сторону	Пояснення очікуваного впливу
Фізичні та юридичні особи, фізичні особи - підприємці	Позитивний	Можливість отримання доступу до якісного українськомовного корпусу даних
Держава	Позитивний	Підвищення популяризації української мови в цифровому просторі, забезпечення її функціонування в сучасних системах штучного інтелекту, а також збереження та поширенню національної культурної та історичної спадщини.

**Міністр цифрової
трансформації України**

Оксана ФЕРЧУК

_____ 2026 р.