КАБІНЕТ МІНІСТРІВ УКРАЇНИ

ПОСТАНОВА

від 11 листопада 2026 р. № 1416

Київ

Деякі питання розвитку українських великих мовних моделей

Кабінет Міністрів України **постановляє**:

1. Затвердити Положення про український корпус даних, що додається.

2. Забезпечити:

1) Міністерству цифрової трансформації протягом трьох місяців з дня набрання чинності цією постановою створення українського корпусу даних;

2) органам виконавчої влади у місячний строк з дня надходження запиту Міністерства цифрової трансформації або державного підприємства "ДІЯ" передачу зазначеному підприємству з дотриманням вимог законодавства про інтелектуальну власність в межах наявних прав наборів знеособлених структурованих та неструктурованих даних, що можуть використовуватися для тренування та тестування великих мовних моделей або українських великих мовних моделей відповідно до Положення, затвердженого цією постановою.

3. Рекомендувати державним органам, державним науковим організаціям та установам, архівним установам, бібліотекам, закладам освіти, а також юридичним та фізичним особам, фізичним особам — підприємцям сприяти створенню українського корпусу даних із дотриманням прав наукових установ, авторів та інших правовласників.

4. Міністерству цифрової трансформації разом з Міністерством освіти і науки за участю Національної академії наук протягом шести місяців після створення українського корпусу даних опрацювати питання щодо подальшого розвитку національної системи мовних ресурсів та за результатами опрацювання подати Кабінетові Міністрів України відповідні пропозиції.

Прем'єр-міністр України **С. КОРЕЦЬКИЙ**

Інд. 81

ЗАТВЕРДЖЕНО
постановою Кабінету Міністрів України
від 11 листопада 2026 р. № 1416

ПОЛОЖЕННЯ
про український корпус даних

1. Це Положення визначає механізм створення, наповнення, підтримання в актуальному стані українського корпусу даних та надання доступу до нього.

2. Метою створення українського корпусу даних є розвиток українських великих мовних моделей шляхом тренування та тестування великих мовних моделей та українських великих мовних моделей.

3. У цьому Положенні терміни вживаються у такому значенні:

велика мовна модель — модель машинного навчання, призначена для обробки природної мови, генерації тексту, аналізу інформації, формування відповідей, перекладу та виконання інших мовних завдань на основі вивчених закономірностей української мови;

користувач українського корпусу даних (далі — користувач) — суб’єкт, що відповідає вимогам, визначеним цим Положенням, з яким укладено ліцензійний договір та який отримав доступ до українського корпусу даних;

набір даних — набір знеособлених структурованих та неструктурованих даних (бази даних, тексти, документи, літературні твори, наукові дослідження, архівні справи, архівні документи тощо), що, зокрема, містять тексти українською мовою та інформацію, яка відображає мовні, культурні та історичні особливості України;

постачальник даних — суб’єкт, який надає набори даних для формування українського корпусу даних;

українська велика мовна модель — велика мовна модель, яка натренована на значних обсягах україномовних текстових даних та виконує мовні завдання на основі вивчених закономірностей української мови з урахуванням культурного та історичного контексту;

український корпус даних (далі — корпус даних) — систематизована та структурована сукупність наборів даних (база (компіляція) даних), що використовуються для тренування та тестування великих мовних моделей та українських великих мовних моделей.

Інші терміни вживаються у значенні, наведеному в Законах України “Про інформацію”, “Про доступ до публічної інформації”, “Про захист персональних даних”, “Про авторське право і суміжні права”, “Про Національний архівний фонд та архівні установи”, “Про бібліотеки і

бібліотечну справу”, Положенні про набори даних, які підлягають оприлюдненню у формі відкритих даних, затвердженому постановою Кабінету Міністрів України від 21 жовтня 2015 р. № 835 (Офіційний вісник України, 2015 р., № 85, ст. 2850).

4. Держателем корпусу даних є Мінцифри.

Створювачем корпусу даних, який за добором та упорядкуванням його складових частин є результатом інтелектуальної діяльності, є державне підприємство “ДІЯ” або його правонаступник (далі — укладач).

Належність авторських прав на складові частини корпусу даних та результати їх технологічного оброблення визначається відповідно до законодавства.

Включення об’єктів права інтелектуальної власності до корпусу даних не змінює належності авторських прав на такі об’єкти та не припиняє майнових прав інтелектуальної власності третіх осіб. Використання складових частин корпусу даних здійснюється в межах прав, наданих їх правовласниками або передбачених законом.

Формування та використання корпусу даних не змінює правового статусу національних мовних лінгвістичних і мовно-знаннєвих ресурсів, зокрема наукових об’єктів, що становлять національне надбання, не передбачає автоматичного переходу прав на них та не змінює визначених законодавством повноважень у сфері стандартизації державної мови. Використання таких ресурсів здійснюється з дотриманням прав їх правовласників та умов надання доступу.

5. Держатель корпусу даних здійснює організаційне забезпечення заходів, пов’язаних із створенням, наповненням, підтриманням в актуальному стані корпусу даних та наданням доступу до нього.

6. Укладач:

створює та наповнює корпус даних, що передбачає збір, накопичення, адаптацію, зміну, поновлення, компіляцію, обробку, очищення, структурування даних та перевірку відповідності отриманих наборів даних вимогам щодо їх знеособлення;

підтримує в актуальному стані корпус даних шляхом включення до його складу нових наборів даних, внесення змін до наборів даних, отриманих від постачальників даних, виключення наборів даних з корпусу даних;

надає та припиняє доступ до корпусу даних користувачам;

проводить навчання щодо роботи із корпусом даних;

затверджує за погодженням з Мінцифри перелік документів для підтвердження юридичними особами або фізичними особами — підприємцями відповідності вимогам, визначеним у пункті 16 цього Положення;

припиняє та поновлює доступ до наборів даних, включених до корпусу даних, у випадках, визначених цим Положенням;

забезпечує зберігання та захист корпусу даних та інформації, що в ньому міститься;

оприлюднює корпус даних на Єдиному державному веб-порталі відкритих даних у частині публічної інформації, а також інформації, правовласниками якої згідно з умовами ліцензійного договору, надано дозвіл на її оприлюднення.

7. Постачальниками даних є державні органи, державні наукові організації та установи, архівні установи, бібліотеки, заклади освіти, а також інші юридичні та фізичні особи, фізичні особи — підприємці.

8. Укладачу надаються набори даних, що містять:

1) виключно публічну інформацію, яка є відкритою, — розпорядниками інформації безоплатно за запитом Мінцифри або укладача;

2) інформацію, яка не є публічною інформацією, — постачальниками даних за їх згодою на підставі договору, укладеного з укладачем, який передбачає право укладача на включення таких даних до корпусу даних та надання користувачам доступу до них у складі корпусу даних відповідно до цього Положення на умовах, визначених відповідним договором.

9. Укладач для наповнення корпусу даних може також отримувати набори даних з офіційних веб-сайтів, Єдиного державного веб-порталу відкритих даних, інших загальнодоступних джерел публічної інформації.

10. Для отримання наборів даних відповідно до підпункту 1 пункту 8 цього Положення у запиті Мінцифри або укладача визначаються склад та спосіб передачі таких наборів даних. Отримані набори даних укладач накопичує, перевіряє, обробляє, очищує, адаптує, структурує, систематизує, компілює, перевіряє на відповідність вимогам щодо їх знеособлення та приводить у відповідність із форматами, придатними для включення та використання у складі корпусу даних.

11. Для отримання наборів даних відповідно до підпункту 2 пункту 8 цього Положення укладач укладає з постачальником даних ліцензійний договір, в якому визначаються склад та спосіб передачі таких наборів даних, а також умови та способи їх використання і обробки укладачем. Отримані набори даних укладач у межах прав, наданих відповідно до такого договору, накопичує, перевіряє, обробляє, очищує, адаптує, структурує, систематизує, компілює, перевіряє на відповідність вимогам щодо їх знеособлення та

приводить до форматів, придатних для включення та використання у складі корпусу даних.

Постачальник даних підтверджує наявність правових підстав для передачі даних, які не є публічною інформацією, та їх використання способами, передбаченими цим Положенням, і несе відповідальність за достовірність наданої інформації.

Якщо використання відповідних даних способами, передбаченими цим Положенням, потребує дозволу правовласника, постачальник даних може передати такі дані укладачу лише у разі наявності прав або повноважень, достатніх для надання користувачам права на їх використання відповідно до цього Положення.

У разі надходження від правовласника претензії, скарги або іншого звернення (далі — скарга) до укладача або Мінцифри постачальник даних надає укладачу документи, що підтверджують правомірність передачі та використання відповідних даних.

12. Постачальники даних передають набори даних укладачу у форматах, придатних для автоматизованого оброблення електронними засобами.

13. За результатами підготовки та обробки наборів даних, отриманих відповідно до пунктів 8 і 9 цього Положення, укладач формує корпус даних шляхом об’єднання, систематизації та структурування таких наборів даних у базу (компіляцію) даних у форматі, придатному для її використання відповідно до мети, визначеної пунктом 2 цього Положення.

14. Сформований укладачем корпус даних у частині добору та упорядкування його складових є результатом інтелектуальної діяльності укладача та належить державі, що не змінює належності прав на складові частини корпусу даних.

15. Укладач реалізує право інтелектуальної власності в частині надання права використання корпусу даних в обсязі, необхідному для:

наповнення, оновлення, підтримання в актуальному стані, зберігання, захисту та забезпечення доступу до нього відповідно до цього Положення;

надання користувачам на безоплатній основі в межах отриманих укладачем прав використання корпусу даних для тренування та тестування великих мовних моделей та українських великих мовних моделей у порядку та на умовах, визначених цим Положенням.

16. Доступ надається до сформованого корпусу даних, а не до окремих його складових частин.

Доступ до корпусу даних надається юридичній особі або фізичній особі — підприємцю, яка звернулася до укладача з метою отримання такого доступу, зобов’язується використовувати корпус даних виключно для

тренування та тестування великих мовних моделей та українських великих мовних моделей (далі — заявник) та одночасно відповідає таким вимогам:

провадить діяльність, пов’язану з розробленням, тренуванням або тестуванням великих мовних моделей та українських великих мовних моделей, та надала документи, що підтверджують наявність такого досвіду;

не провадить свою господарську діяльність на тимчасово окупованих Російською Федерацією територіях України та територіях активних бойових дій, включених до переліку територій, на яких ведуться (велися) бойові дії або тимчасово окупованих Російською Федерацією, затвердженого в установленому порядку, та для яких не визначено дати завершення бойових дій або тимчасової окупації;

не провадить господарську діяльність на території Російської Федерації, Республіки Білорусь та/або держав, які входять до митних або воєнних союзів із такими державами;

не належить до юридичних або фізичних осіб, до яких застосовуються спеціальні економічні та інші обмежувальні заходи (санкції) відповідно до рішень Ради національної безпеки і оборони України, введених у дію указами Президента України щодо застосування персональних (спеціальних) економічних та інших обмежувальних заходів (санкцій) згідно із Законом України “Про санкції”;

не є фінансовою установою, інвестиційним фондом, недержавним пенсійним фондом, професійним учасником ринків капіталу та організованих товарних ринків;

кінцевими бенефіціарними власниками якої не є громадяни держави, що здійснює збройну агресію проти України (крім громадян такої держави, яким надано статус учасника бойових дій після 14 квітня 2014 р.), та/або особи, місцем постійного проживання (перебування, реєстрації) яких є держава, що здійснює збройну агресію проти України;

не є фізичною особою підприємцем, що має громадянство держави, визнаної відповідно до закону державою-агресором або державою-окупантом, крім тих, що на законних підставах постійно перебувають в Україні;

не є юридичною особою, утвореною відповідно до законодавства держави, що визнана відповідно до закону державою-агресором або державою-окупантом;

не є юридичною особою, в якій держава, визнана відповідно до закону державою-агресором або державою-окупантом, юридичною особою, утвореною відповідно до законодавства такої держави, або громадянином такої держави, крім того, що постійно на законних підставах перебуває на території України, є учасником (акціонером), частка якого в статутному

капіталі з урахуванням відносин контролю становить 10 або більше відсотків, або членом, або кінцевим бенефіціарним власником.

17. Для отримання доступу до корпусу даних заявник надсилає укладачу повідомлення про намір отримати доступ до корпусу даних (далі — повідомлення про намір), до якого додаються документи, що підтверджують відповідність заявника вимогам, визначеним пунктом 16 цього Положення, відповідно до переліку, затвердженого укладачем за погодженням з Мінцифри.

Повідомлення про намір повинно містити найменування юридичної особи або прізвище, власне ім'я, по батькові (за наявності) фізичної особи — підприємця, мету отримання доступу до корпусу даних, відомості про діяльність заявника, пов'язану з розробленням, тренуванням або тестуванням великих мовних моделей та українських великих мовних моделей, адресу електронної пошти, а також підтвердження відповідності заявника вимогам, визначеним пунктом 16 цього Положення.

Повідомлення про намір та документи, що додаються до нього, надсилаються в електронній формі на адресу електронної пошти, визначену укладачем та оприлюднену на його офіційному веб-сайті або офіційному веб-сайті Мінцифри.

Укладач протягом 10 робочих днів з дня надходження повідомлення про намір та документів, що додаються до нього, здійснює їх розгляд та проводить перевірку відповідності заявника вимогам, визначеним пунктом 16 цього Положення.

У разі виявлення недоліків у поданому повідомленні про намір або документах, що додаються до нього, укладач протягом п'яти робочих днів з дня їх надходження інформує заявника про виявлені недоліки шляхом надсилання відповідного повідомлення на адресу електронної пошти, зазначену в повідомленні про намір, із зазначенням способу їх усунення.

Заявник протягом 10 робочих днів з дня отримання повідомлення про виявлені недоліки має право їх усунути. Перебіг строку розгляду повідомлення про намір зупиняється з дня надсилання укладачем повідомлення про виявлені недоліки та поновлюється з дня надходження уточнених відомостей або документів.

За результатами розгляду повідомлення про намір та документів, що додаються до нього, укладач приймає рішення про відповідність або невідповідність заявника вимогам, визначеним пунктом 16 цього Положення.

Підставами для прийняття рішення про невідповідність заявника вимогам, визначеним пунктом 16 цього Положення, є невідповідність заявника хоча б одній із зазначених вимог, виявлення недостовірних відомостей у поданому повідомленні про намір або документах, що

додаються до нього, або неусунення заявником виявлених недоліків у встановлений строк.

Укладач протягом трьох робочих днів з дня прийняття відповідного рішення інформує заявника про результати розгляду повідомлення про намір шляхом надсилання повідомлення на адресу електронної пошти, зазначену заявником, із зазначенням підстав для прийняття рішення про невідповідність заявника вимогам, визначеним пунктом 16 цього Положення, у разі прийняття такого рішення.

Прийняття рішення про невідповідність заявника вимогам, визначеним пунктом 16 цього Положення, не позбавляє заявника права повторно надіслати повідомлення про намір після усунення обставин, що стали підставою для прийняття такого рішення.

Заявник забезпечує достовірність відомостей та документів, поданих для отримання доступу до корпусу даних.

У разі зміни відомостей, що впливають на відповідність користувача вимогам, визначеним пунктом 16 цього Положення, користувач зобов'язаний протягом десяти робочих днів з дня виникнення таких змін повідомити про них укладачу у спосіб, передбачений для надсилання повідомлення про намір.

У разі отримання інформації про можливу невідповідність користувача вимогам, визначеним пунктом 16 цього Положення, укладач проводить перевірку такої інформації.

У разі встановлення за результатами перевірки невідповідності користувача хоча б одній з вимог, визначених пунктом 16 цього Положення, укладач приймає рішення про припинення доступу такого користувача до корпусу даних та протягом трьох робочих днів з дня прийняття відповідного рішення інформує його про це шляхом надсилання повідомлення на адресу електронної пошти, зазначену в повідомленні про намір, із зазначенням підстав для прийняття такого рішення.

Рішення укладача про невідповідність заявника вимогам, визначеним пунктом 16 цього Положення, може бути оскаржене до Мінцифри протягом 10 робочих днів з дня отримання заявником повідомлення про прийняття такого рішення.

18. У разі прийняття укладачем рішення про відповідність заявника вимогам, визначеним пунктом 16 цього Положення, укладач та заявник укладають ліцензійний договір.

Доступ до корпусу даних надається після укладення ліцензійного договору між укладачем та заявником.

Після укладення ліцензійного договору заявник набуває статусу користувача.

Доступ до корпусу даних є безоплатним.

19. У ліцензійному договорі обов’язково визначаються умови надання та припинення доступу до корпусу даних та його використання користувачем, зокрема щодо використання корпусу даних виключно для тренування, тестування великих мовних моделей та українських великих мовних моделей і заборони використання корпусу даних в інший спосіб.

20. Користувач може використовувати корпус даних виключно для тренування, тестування великих мовних моделей та українських великих мовних моделей.

У разі виявлення укладачем випадку використання користувачем корпусу даних в інший спосіб, ніж визначений у цьому пункті, укладач негайно вживає заходів до припинення доступу такого користувача до корпусу даних.

21. У разі необхідності оновлення корпусу даних укладач звертається до Мінцифри з пропозицією щодо його актуалізації. У такому зверненні укладач зазначає набори даних, що потребують оновлення.

22. Набори даних, необхідні для актуалізації корпусу даних, надаються постачальниками даних у порядку, встановленому пунктами 8—11 цього Положення.

23. Після отримання відповідних наборів даних такі набори даних обробляються укладачем у порядку, встановленому цим Положенням, та використовуються для актуалізації корпусу даних шляхом включення нових наборів даних, оновлення наборів даних, включених до його складу, та/або виключення наборів даних, що втратили актуальність.

24. У разі надходження обґрунтованої скарги правовласника щодо можливого порушення авторського права у зв’язку з включенням окремого набору даних до корпусу даних або його використанням укладач не пізніше одного робочого дня з дати отримання відповідної скарги тимчасово обмежує доступ до такого набору даних до завершення розгляду скарги.

Укладач не пізніше одного робочого дня з дня отримання скарги, передбаченої цим пунктом, повідомляє Мінцифри про надходження такої скарги. У разі надходження зазначеної скарги до Мінцифри така скарга протягом одного робочого дня з дня її отримання надсилається укладачу, інформування Мінцифри про надходження скарги в такому випадку не здійснюється.

Скарга повинна містити відомості про правовласника, набір даних, щодо якого подається скарга, опис обставин можливого порушення авторського права та обґрунтування відповідних вимог. До скарги додаються наявні документи, що підтверджують належність авторських прав на відповідні об’єкти.

Укладач забезпечує розгляд такої скарги протягом 10 робочих днів з дня її отримання та інформує Мінцифри про результати її розгляду з пропозиціями щодо подальшого використання відповідного набору даних.

За результатами розгляду скарги укладач приймає рішення про поновлення доступу до відповідного набору даних або його виключення із складу корпусу даних.

Поновлення доступу до набору даних, доступ до якого тимчасово обмежено відповідно до цього пункту, здійснюється за погодженням з Мінцифри. Виключення набору даних із складу корпусу даних є підставою для остаточного припинення доступу до такого набору даних у складі корпусу даних.

Укладач повідомляє правовласнику про результати розгляду скарги протягом трьох робочих днів з дня прийняття відповідного рішення.

25. Створення, наповнення, підтримання в актуальному стані українського корпусу даних та надання доступу до нього здійснюється з дотриманням вимог законодавства про захист персональних даних та інформації з обмеженим доступом та авторське право і суміжні права.

_____