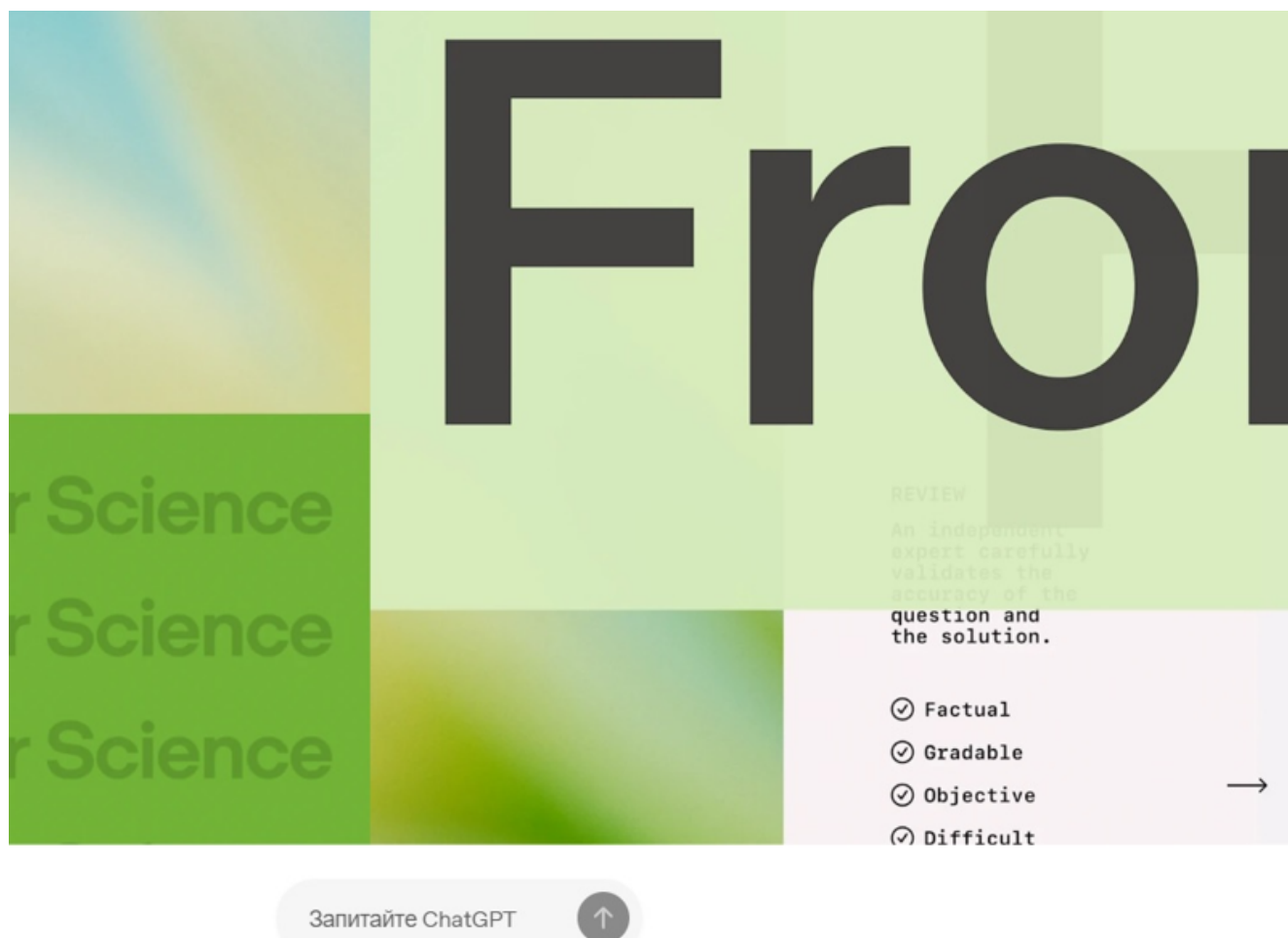


ОЦІНКА ЗДАТНОСТІ ШТУЧНОГО ІНТЕЛЕКТУ ВИКОНУВАТИ НАУКОВО-ДОСЛІДНІ ЗАВДАННЯ



На сайті OpenAI опублікована інформація «Оцінка здатності штучного інтелекту виконувати науково-дослідні завдання».

У ній представлено новий бенчмарк FrontierScience, який оцінює можливості штучного інтелекту для проведення наукових міркувань експертного рівня у фізиці, хімії та біології. Окрім роботи з фактами вчені генерують гіпотези, перевіряють та уточнюють їх, а також синтезують ідеї. Оскільки моделі ШІ стають більш потужними, центральне питання полягає в тому, чи можуть вони міркувати, щоб зробити свій внесок у наукові дослідження. Упродовж останнього року моделі ГШІ досягли важливих успіхів (перемога на Міжнародній математичній олімпіаді та Міжнародній олімпіаді з інформатики). Разом із тим, найпотужніші моделі, такі як GPT-4, суттєво пришвидшують реальні наукові робочі процеси. Дослідники використовують ці системи для таких завдань, як пошук літератури з різних дисциплін та мов програмування, робота зі складними математичними доказами. Цей прогрес задокументовано в статті «Ранні експерименти з

прискорення науки з GPT^{UA}5», опублікованій у листопаді 2025 року. У ній представлені докази того, що GPT^{UA}5 може помітно пришвидшити наукові робочі процеси. Оскільки можливості міркування та знань моделей продовжують масштабуватися, складніші тести будуть важливими для вимірювання та прогнозування здатності моделей пришвидшувати наукові дослідження. Попередні наукові тести значною мірою зосереджені на питаннях з множинним вибором, є перенасиченими або не зосереджені центрально на науці. Щоб подолати цю проблему, FrontierScience створений для вимірювання наукових можливостей експертного рівня. FrontierScience написаний та перевірений експертами з фізики, хімії та біології і складається з сотень питань, розроблених таким чином, щоб бути складними, оригінальними й змістовними. Він включає два треки питань: Olympiad (вимірює можливості наукового мислення в стилі олімпіади) та Research (вимірює здібності до реальних наукових досліджень). Згідно з початковими оцінками, GPT^{UA}5.2 є найкращою моделлю на FrontierScience-Olympiad (77%) та Research (25%), випереджаючи інші моделі у галузі досліджень. Спостерігається значний прогрес у вирішенні питань експертного рівня, залишаючи при цьому простір для подальшого прогресу, особливо у завданнях відкритого дослідницького типу. Для науковців це означає, що сучасні моделі вже можуть частково підтримувати дослідження, які передбачають структуроване мислення, але ще потрібно попрацювати над покращенням їхньої здатності здійснювати відкрите мислення. Отримані результати узгоджуються з тим, як науковці вже використовують сучасні моделі: для прискорення дослідницьких робочих процесів, покладаючись на людське судження для формулювання та перевірки проблем, а також для все більшого дослідження ідей та зв'язків, розкриття яких в іншому випадку зайняло б набагато більше часу, включаючи, в деяких випадках, внесення нових висновків, які потім оцінюють та перевіряють експерти. Найважливішим орієнтиром для наукових можливостей ШІ є нові відкриття, які він допомагає створювати: саме вони зрештою мають значення для науки та суспільства. FrontierScience дає орієнтир для наукових міркувань експертного рівня, дозволяючи тестувати моделі на стандартизованому наборі питань, бачити, де вони успішні, а де невдалі, а також визначати, де потрібно їх покращити. FrontierScience є вузькою програмою та має обмеження в ключових аспектах (зосередження на обмежених, написаних експертами проблемах) і не охоплює все, що роблять вчені у своїй повсякденній роботі. Ця галузь потребує більш складних, оригінальних та змістовних наукових орієнтирів, а FrontierScience робить крок вперед у цьому напрямку.

Детальніше: <https://qrpage.net/qr/RWrYl>, <https://qrpage.net/qr/v3IEy>

Фото: скріншот

#НРАТ_Усі_новини #НРАТ_ШтучнийІнтелект #НРАТ_ВідкритаНаука
#НРАТ_Науковцям_новини #НРАТ_Освіт'ям_новини
#НРАТ_Інноваторам_новини #НРАТ_Бібліотекарям

2025-12-18

**Інформація з офіційного вебпорталу Національного репозитарію
академічних текстів**