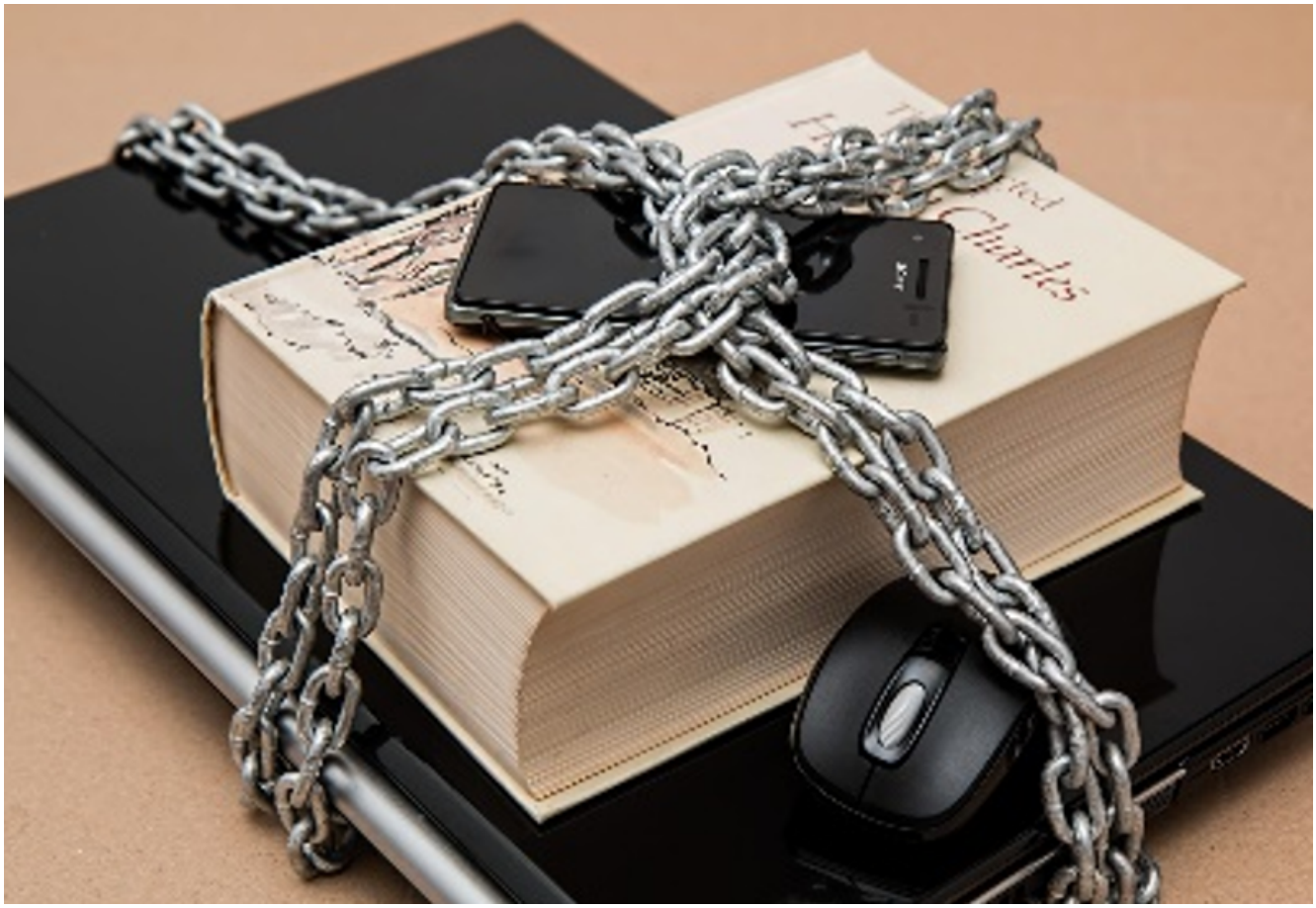
https://nrat.ukrintei.ua/gumanitarni-nauky-povynni-doluchytys-do-kontrolyu-czenzury-shtuchnogo-intelektu/

ГУМАНІТАРНІ НАУКИ ПОВИННІ ДОЛУЧИТИСЬ ДО КОНТРОЛЮ «ЦЕНЗУРИ» ШТУЧНОГО ІНТЕЛЕКТУ



На сайті Times Higher Education опублікована стаття Лорни Ваддінгтон та Річарда де Блак'єр-Кларксона «Гуманітарні науки повинні відігравати певну роль у контролі за «цензурою» штучного інтелекту».

Автори зазначають, що певні обмеження в ШІ-моделях необхідні, але не можна припуститись спотворення реальної картини. Відомий інцидент, коли у травні 2025 року чат-бот Grok3 почав створювати повідомлення про насильство щодо білих людей у Південній Африці або «геноцид білих». Компанія звинуватила у цьому «несанкціоновану модифікацію» свого програмного забезпечення, яка «порушувала внутрішню політику та основні цінності xAI». Але це був не єдиний випадок із ШІ. Лорна та Річард описують очевидні наслідки для науковців, які використовують штучний інтелект у своїх дослідженнях і наголошують, що подібні інциденти роблять важливим питанням саме модерацію контенту. З одного боку, ШІ обіцяє розширені аналітичні можливості, що дозволяють нам обробляти величезні обсяги даних з безпрецедентною швидкістю та точністю. З іншого боку, його непрозорі та часом довільні механізми фільтрації загрожують

створити штучні прогалини в нашому розумінні світу, його історії та нашого місця в ньому. Надзвичайно важко приймати збалансовані рішення у цій сфері, але вони потрібні. Саме тому компаніям, що займаються штучним інтелектом, потрібен експертний академічний внесок. Ретельно зібрана команда з широкого кола дисциплін з гуманітарним досвідом та усталеними методологічними традиціями буде при її залученні до такої роботи доволі ефективною. Зі свого боку, розробники повинні публікувати чітку документацію, в якій викладено критерії, що використовуються для фільтрації контенту, а також положення про їх скасування перевіреними дослідниками. Тримати людей в курсі подій життєво важливо, щоб максимально використати переваги штучного інтелекту та мінімізувати його ризики.

Детальніше: https://qrpage.net/qr/TZtvp

Фото: pixabay.com

#НРАТ_Усі_новини #НРАТ_ШтучнийІнтелект #НРАТ_Науковцям_новини #НРАТ_Освітянам_новини #НРАТ_Інноваторам_новини

2025-07-24

Інформація з офіційного вебпорталу Національного репозитарію академічних текстів