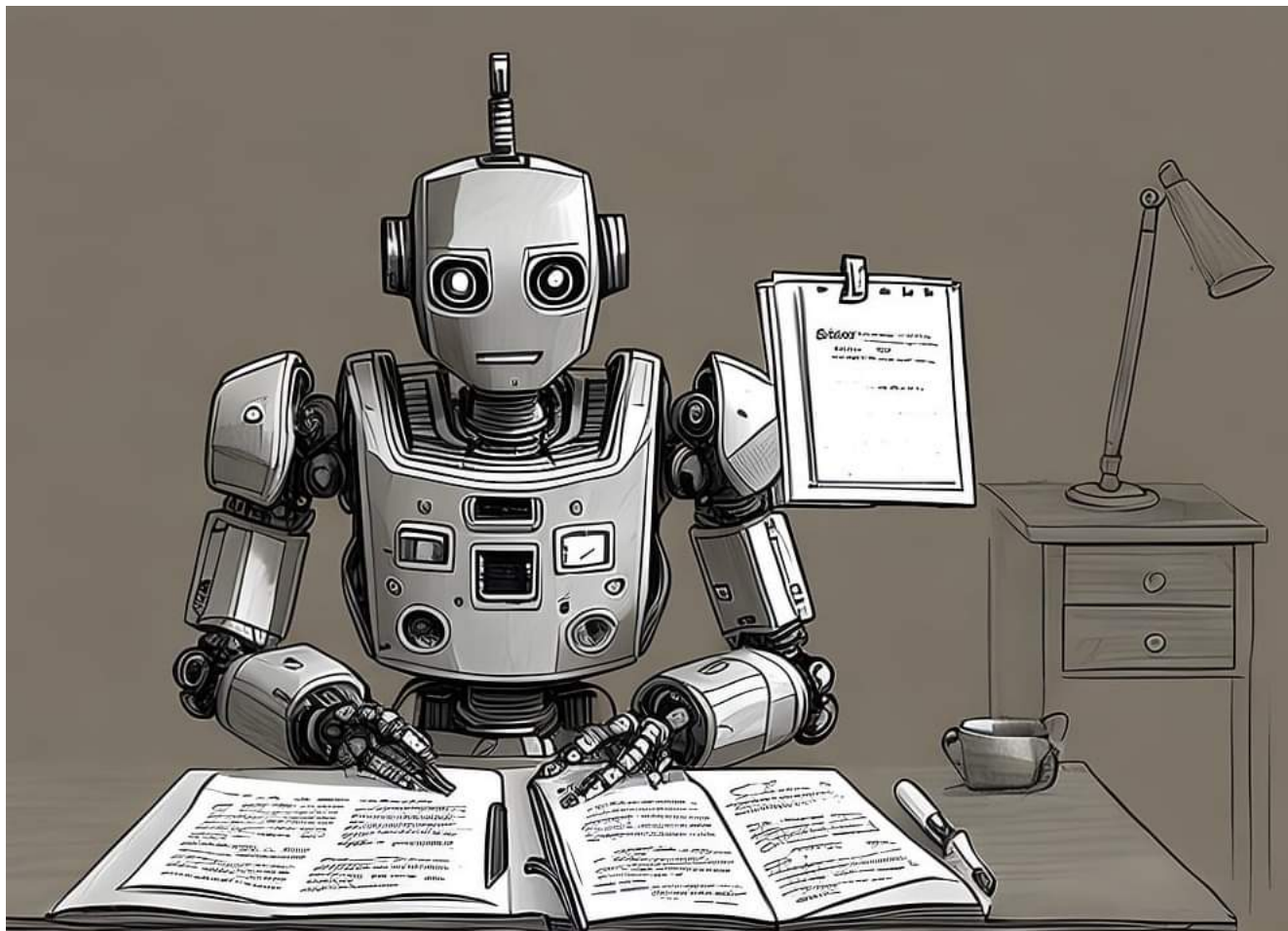


ДЕТЕКТОРИ ТЕКСТУ ШІ НЕ ПРАЦЮЮТЬ



На сайті Times Higher Education опублікована стаття Тома Вільямса «Детектори тексту ШІ не працюють».

Чи є регулювання відповіддю?». Автор розмірковує над питанням: оскільки інструменти, розроблені для викорінення неправомірних дій з інструментами ШІ виявились неточними, чи потрібно більш суворе регулювання використання генеративного штучного інтелекту в академічному середовищі? З появою ШІ університети в усьому світі підтримали використання відповідних детекторів задля протидії недобросовісним діям студентів під час навчання. Але наразі стає все більше доказів того, що «доступні інструменти виявлення не є ані точними, ані надійними та в основному схильні класифікувати результати як написані людиною, а не виявляти текст, згенерований штучним інтелектом». Дослідження, проведене вченими з Університету Меріленда, підтвердило побоювання з приводу неточності роботи детекторів: студенти можуть легко їх перехитрити, використовуючи інструменти перефразування для переписування тексту, створеного за допомогою великих мовних моделей. Один з авторів цього дослідження, Сохейл Фейзі, доцент кафедри комп'ютерних наук, сказав, що недоліки в інструментах уже проявили себе і

мали негативний вплив (несправедливе звинувачення студентів у неправомірній поведінці). «Зростає визнання того, що нам необхідна можливість розрізняти текст, написаний штучним інтелектом, і текст, написаний людиною з низки етичних та юридичних причин. У довгостроковій перспективі на користь кожного знати, чи створено щось ШІ чи ні», вважає професор Дрейпер. Модель детектора, створена OpenAI, нещодавно була відкладена, що багато хто розцінив як доказ того, що виявлення ГШІ неможливе. Turnitin, чий детектор загалом отримав більш високі оцінки, ніж більшість інших та був використаний 65 мільйонів разів, теж не став непогрішимим. Нещодавно у США великими компаніями, що працюють над великими мовними моделями, було прийнято добровільне зобов'язання щодо розробки «надійних технічних механізмів, які гарантують, що користувачі знають, коли контент є штучним інтелектом». «Дехто каже, що виявлення ніколи не йтиме в ногу з часом. Це вірно, коли незалежна компанія намагається передбачити, що відбудеться далі, але коли у вас є зобов'язання створити засоби виявлення ГШІ від самих компаній-розробників штучного інтелекту, ви перебуваєте на набагато сильнішій позиції».

Детальніше: <https://is.gd/1W6VSS>,
<https://arxiv.org/ftp/arxiv/papers/2306/2306.15666.pdf>,
<https://arxiv.org/pdf/2303.11156.pdf>

Фото: pixabay.com

#НРАТ_Усі_новини #НРАТ_АкадемДоброчесність #НРАТ_ШтучнийІнтелект
#НРАТ_Науковцям_новини #НРАТ_Освітянам_новини
2023-08-29

Інформація з офіційного вебпорталу Національного репозитарію академічних текстів