
Detecció de contingut fet per IA - Algorismes

Autor:

Data de publicació: 31-08-2025

Els algorismes de detecció de contingut fet per intel·ligència artificial (IA) pretenen detectar si un text, o fins i tot una imatge de vídeo o un àudio han estat produïts per éssers humans o per intel·ligència artificial.

A partir del 2023, han sortit al mercat diferents aplicacions que pretenen detectar si el text ha estat creat per intel·ligència artificial i, de fet s'utilitzen als col·legis i universitats per poder detectar els treballs no originals dels estudiants. Aquest ús és controvertit, ja que les aplicacions emprades no són fiables i poden ser utilitzades incorrectament per professors i catedràtics.[1]

Els algorismes de detecció de contingut fet per intel·ligència artificial (IA) pretenen detectar si un text, o fins i tot una imatge de vídeo o un àudio han estat produïts per éssers humans o per intel·ligència artificial.

A partir del 2023, han sortit al mercat diferents aplicacions que pretenen detectar si el text ha estat creat per intel·ligència artificial i, de fet s'utilitzen als col·legis i universitats per poder detectar els treballs no originals dels estudiants. Aquest ús és controvertit, ja que les aplicacions emprades no són fiables i poden ser utilitzades incorrectament per professors i catedràtics.[1]

Detecció de text

En el cas de document de text, això es fa habitualment per prevenir el suposat plagi, detectant la repetició de paraules com a senyals reveladors que un text ha estat generat per IA (incloent les al·lucinacions d'IA). Sovint són utilitzats pels professors que investiguen els seus alumnes, normalment de manera ad hoc . Després del llançament de ChatGPT i programes similars de generació de textos d'IA, molts centres educatius han dictat polítiques contra l'ús de la IA per part dels estudiants.[2] Els programes de detecció de text d'IA també els utilitzen els que avaluen els sol·licitants de feina, així com els motors de cerca en línia.[3]

Els detectors actuals no són fiables i han marcat incorrectament treball d'humans com a originari de la IA[4][5][6] mentre no detecten el treball generat per la IA en altres casos.[7] La revista MIT Technology Review va dir que la tecnologia "va lluitar per recollir text generat per ChatGPT que havia estat lleugerament reordenat pels humans i ofuscat per una eina de parafraseig".[8] També s'ha demostrat que el programes de detecció de text d'IA discrimina als parlants no nadius d'anglès.[3]

El maig de 2023, un professor de la Texas University–Commerce va utilitzar ChatGPT per detectar si el treball dels seus estudiants estava escrit per ell, cosa que ChatGPT va dir que així era. Com a tal, va amenaçar amb suspendre la classe, tot i que ChatGPT no va poder detectar quins eren els escrits generats per IA.[9] No es va impedir a cap estudiant graduar-se per raó s'aquest problema, i tots els estudiants menys un, que va reconèixer haver-ho fet, van ser exonerats d'haver utilitzat ChatGPT en el seu treball.[10] L'abril de 2023, la Universitat de Cambridge i altres membres del grup d'universitats Russell van desestimar l'eina de detecció de textos d'IA de Turnitin, argumentant que no era fiable.[11]

Existeixen una sèrie de suposats programes de detecció d'imatges d'IA per a intentar detectar imatges generades artificialment (per exemple, aquelles originades a Midjourney), però no són fiables.[12] Altres programes afirmen identificar deepfakes de vídeo i àudio, però aquesta tecnologia tampoc és fiable.[13]

Referències

- «AI Content Detector Checks GPT-4, ChatGPT, Bard, & More». [Consulta: 23 agost 2023].
- «AI-assisted plagiarism? ChatGPT bot says it has an answer for that». The Guardian, 31-12-2022.
- «Programs to detect AI discriminate against non-native English speakers, shows study». The Guardian, 10-07-2023.
- «Detecting AI may be impossible. That's a big problem for teachers.». , 02-06-2023.
- «There's a Problem With That App That Detects GPT-Written Text: It's Not Very Accurate». , 09-01-2023.
- «We tested a new ChatGPT-detector for teachers. It flagged an innocent student.». , 01-04-2023.
- «ChatGPT maker OpenAI releases 'not fully reliable' tool to detect AI generated content». , 01-02-2023.
- «AI-text detection tools are really easy to fool». , 07-07-2023.
- «A professor accused his class of using ChatGPT, putting diplomas in jeopardy». , 18-05-2023.
- «College instructor put on blast for accusing students of using ChatGPT». , 18-05-2023.
- «Universities express doubt over tool to detect AI-powered plagiarism». , 03-04-2023.
- ; Hsu, Tiffany «How Easy Is It to Fool A.I.-Detection Tools?». The New York Times, 28-06-2023.
- ; Myers, Steven Lee «Another Side of the A.I. Boom: Detecting What A.I. Makes». The New York Times, 18-05-2023.