

Albert Sabater: “Els accidents més greus en IA són de caràcter quotidià”

ONA VILÀ I COLOMER

La intel·ligència artificial (IA) ha avançat moltíssim en els últims 70 anys. Dels primers sistemes d'estadística avançada que utilitzava l'INE per comptar ràpidament la població en un cens hem passat a la possibilitat de crear continguts de cap i de nou. Ara bé, és realment intel·ligent la IA? És lícit fer-la servir? Quins límits li hem de posar? En parlem amb **Albert Sabater**, professor Serra Húnter de sociologia i director de l'Observatori d'Ètica en Intel·ligència Artificial de Catalunya. Una càtedra de la Universitat de Girona que analitza les conseqüències ètiques i socials de la implantació cada vegada més generalitzada de la IA.

A què us referiu quan parleu d'ètica en la intel·ligència artificial?

No ho hem d'entendre com una capacitat que té la IA per perdre decisions ètiques, no existeix una ètica de la IA, existeixen els usos ètics de la IA. És a dir, som nosaltres que utilitzem una sèrie d'eines, i l'ús que en fem pot ser més o menys transparent, més o menys responsable, més o menys just. En el camp de la IA, ens preocupen les conseqüències en termes de privacitat, d'autonomia, de justícia, de transparència, és a dir, de defensa de drets.

Recentment, s'ha arribat a un acord polític per tirar endavant la primera llei europea d'intel·ligència artificial, què suposa aquest pas?

Aquest acord té dos elements claus que fan que s'avanci cap a un marc normatiu: la transparència i la responsabilitat. La transparència té a veure amb ser més transparents respecte a les eines que s'utilitzen per a crear més confiança, i perquè no tan sols se'n beneficiïn les empreses, sinó també el conjunt de persones que les utilitzen com a consumidors o usuaris finals. Pel que fa a la responsabilitat, no és exclusiva de les persones que desenvolupen sistemes d'IA, sinó també de l'administració, perquè és aquesta qui regula quin tipus d'IA es considera inacceptable, d'alt risc, de baix risc, etc. I alhora, els usuaris també tenim certa responsabilitat. Aquest acord deixa molt clar que cada vegada es demanarà més informació per saber quina és la responsabilitat humana que hi ha al darrere de la presa de decisions més automatitzada dels sistemes d'IA, i en qualsevol cas, la responsabilitat última i legal serà de les persones.

“La responsabilitat última i legal en l'àmbit de la IA serà de les persones”

Hi ha irresponsabilitats en la IA? Es treuen al mercat sistemes poc provats?

Massa habitualment trobem sistemes d'IA que han estat més provats en uns contextos que en uns altres, i quan surten del laboratori, alguna situació evitable no s'ha evitat perquè no s'ha fet prou recerca. La particularitat dels sistemes d'IA és que són molt ràpids i eficients, per tant, si una cosa la fan malament, també la fan molt ràpid i molt malament. A més, els que tenim en l'actualitat són molt complexos i el paper de les empreses que els desenvolupen ha canviat. Ara ja no són els Estats qui porten la iniciativa en la nova revolució tecnològica, sinó les empreses que hi inverteixen, i hi ha un factor de secret industrial que impedeix el principi de transparència que comentàvem abans. Això és molt important també en termes de responsabilitat, perquè

estem davant d'empreses que en molts casos no tenen seu a Catalunya, ni a Espanya, ni a Europa i, per tant, les seves coordenades d'acció són diferents de les que nosaltres podem controlar.

Un dels punts que va dificultar més aquest primer acord polític va ser l'ús de la identificació biomètrica, pel dilema d'haver de triar entre seguretat i privacitat. Què ha de tenir més pes en una decisió així?

Aquí ens hem equivocat, perquè hem fet valer l'argument que la seguretat és més important, quan en realitat aquests sistemes no ens garanteixen cap seguretat. I en cap cas ens garanteixen un manteniment dels nostres drets en termes de vigilància a la via pública. Per exemple, el Regne Unit té una tradició de vigilància molt gran, i això no ha fet que reduís el crim ordinari ni el crim organitzat. La vigilància condiona moltíssim el comportament individual i social, i augmenta les desigualtats, ja que es posen càmeres allà on es pressuposa que hi haurà un problema, però potser el problema no és només allà.

“Els sistemes de vigilància biomètrica no ens garanteixen cap seguretat”

La privacitat, però, en l'àmbit de la IA va molt més enllà dels sistemes de vigilància biomètrica. L'última generació de sistemes d'IA funcionava amb l'aprenentatge automàtic a partir de grans volums de dades. Això suposava un problema a través de l'economia de vigilància, perquè com més dades té un sistema millor, i aquestes dades es recullen a través de la nostra navegació a Internet i de les xarxes gràcies al nostre consentiment, sovint inconscient. Això es va limitar quan va aparèixer la llei de protecció de dades. Tu pots demanar que esborrin totes les teves dades i que no en recullin més. El problema és que amb l'arribada de la IA generativa, hi ha empreses, que sense cap mena de consentiment han agafat informació de bases de dades d'Internet i les han utilitzat sense consentiment per entrenar els seus models i generar nous textos, imatges, sons, vídeos, etc. Entre els acadèmics es coneix com un dels robatoris més grans de la història.

Per tant, la IA generativa ha canviat el paradigma.

Hem passat d'un graó en què estàvem vigilats a un graó en què a més treballem per aquestes companyies per tal de crear una nova història. Els sistemes d'IA generativa com OpenAI, Gemin AI o Mistral AI permeten fer coses perquè prèviament nosaltres ja les hem fet, no poden operar sense tota aquesta informació. I ara estem pagant per uns serveis després d'haver-los cedit gratuïtament tota la informació. Però això canviarà, l'acord polític deixa molt clar que s'hauran de publicar totes les dades que s'han utilitzat i la seva procedència. Tot i això, no estem davant d'un canvi únicament tecnològic, sinó també social. Aquestes eines afecten els nostres drets i llibertats, el valor de les coses i la nostra economia. Les decisions que prenen grans companyies com Microsoft, Amazon o Appel afecten la nostra vida quotidiana i, per tant, hi ha d'haver una sèrie d'elements reguladors que garanteixin que ho fan de manera beneficiosa i no perjudicial.

Aleshores, és un perill l'auge de la IA?

No, de cap de les maneres. Tot i que sempre hi ha una possibilitat que les coses vagin molt malament si no s'inverteix prou en la supervisió i els mecanismes de seguretat. El problema és quan deixa d'haver-hi supervisió humana, llavors hi poden haver accidents. I els accidents més greus no són de caràcter apocalíptic, són més quotidians. Per exemple, que en àmbits com la

medicina o els serveis socials hi hagi biaixos i les eines estigmatitzin algunes poblacions. El perill és no tenir en compte aquestes consideracions ètiques que en molts casos ja formen part de la normativa.

“Les decisions que prenen Microsoft, Amazon o Appel afecten la nostra vida quotidiana”

La IA generativa també ha comportat problemes de veracitat, hi ha qui l'utilitza amb finalitats il·lícites o per crear continguts falsos.

És molt llaminer poder fer veure que alguna cosa és real quan no ho és, i amb la IA passarà cada vegada més. Cal educació i cultura crítica en l'àmbit públic. No hem d'anar accelerats i pensar que tot ens ho hem de creure d'entrada. L'acord polític proposa etiquetar tot el contingut fet amb IA. Però potser la gent no ho farà per una qüestió de com valoren aquest contingut, especialment si la IA el fa millor que les persones. Tot i això, crec que tant tècnicament com socialment, cada vegada hi haurà més consciència que no és acceptable fer passar els resultats de la IA pels resultats de les persones. Una cosa és que la tecnologia ens ajudi i augmenti les nostres possibilitats i una altra és que la tecnologia ho faci absolutament tot i nosaltres ens n'atribuïm l'autoria, això és un frau.

Després d'aquest gran auge, quina tendència creus que seguirà ara la IA?

Ara estem en una fase d'enlluernament, però veurem fins a quin punt es consolida. No saber quines dades s'estan utilitzant i si estan transgredint els drets d'autor fa que les empreses siguin una mica reticents a utilitzar-les per segons què, perquè presenten problemes, sobretot de privacitat. Tot i això, el debat actual i que mou més la inversió és la idea de la IA com una rèplica de la nostra intel·ligència, és a dir que les màquines tinguin una capacitat tan sofisticada com la nostra de fer les coses, que tinguin emocions, empatia, etc. I és molt probable que d'aquí a pocs anys, quan veiem que no hem arribat a aquesta fase de superintel·ligència, el volum d'inversió en IA baixi.

És un viable arribar a aquest estadi en què les màquines tinguin superintel·ligència?

No. De fet, encara tenim un debat de què és la intel·ligència, perquè la intel·ligència no és individual, sinó que és col·lectiva i, per tant, humana. Tenint en compte el que s'està fent actualment per entrenar els sistemes d'IA, no crec que pugui donar una superintel·ligència. No podem fer sistemes que siguin autònoms i intel·ligents com nosaltres únicament amb dades, perquè aquestes tenen uns límits. Una altra cosa seria que s'invertís en IA d'una altra manera, però això ja donaria per una altra entrevista.