



Eugenia Fuenmayor

DIRECTORA CIENTÍFICA DE L'ÀREA DIGITAL D'EURECAT

Les discriminacions a les xarxes, majoritàriament de gènere i raça, provocades per algoritmes computacionals s'estan agreujant amb la integració de la intel·ligència artificial en els sistemes, una situació que pot empitjorar amb el desenvolupament de la modalitat generativa d'aquesta tecnologia **Per M. Sardà**

“L'ètica no pot quedar al marge de l'evolució tecnològica imparable”

Els algoritmes computacionals són una sèrie d'instruccions sistemàtiques i prèviament definides que s'utilitzen en múltiples tasques. Una d'aquestes tasques és optimitzar els resultats de les consultes a internet, funció que realitzen els algoritmes de recerca i ordenació. Tanmateix, aquests algoritmes reflecteixen els valors de qui els codifica, de qui desenvolupa les instruccions, i aquí és on es creen biaixos que discriminen per gènere, raça o llengua.

Quina és la finalitat de les investigacions sobre els biaixos a la xarxa que du a terme l'àrea digital del centre tecnològic Eurecat?

La finalitat última és aconseguir un internet amb qualitat d'informació pel que fa al *big data*. Treballem en la identificació de discursos d'odi i de notícies falses, i proposem solucions per fer que internet sigui tan transparent com sigui possible. Fem vigilància tecnològica continuada perquè en cada moment la indústria pugui aplicar les innovacions.

Som un pont entre la universitat i l'empresa.

També analitzen l'ús de la intel·ligència artificial?

Treballem en pro d'una intel·ligència artificial justa i transparent. Identifiquem els biaixos en algoritmes d'entrenament, el risc de privacitat en les dades, i apliquem solucions que els minimitzen o els eliminen.

Quines investigacions desenvolupen?

Treballem en projectes d'investigació aplicada, en consorcis públics i també en d'altres finançats internament.

Són abundants els biaixos a les xarxes?

Les discriminacions les trobem en molts àmbits. Per exemple, en els sistemes d'intel·ligència artificial de Google o Amazon. Recordo una proposta de feina que va presentar en una ocasió Google, en què es prioritzaven els homes davant les dones, i Amazon també ho ha fet, en aquest sentit, pel que fa a ocupacions tecnològiques. Això respecte a ofertes de feina, però tenim exemples molt més quotidians que



La nostra gran preocupació és la privacitat de les dades que es donen als algoritmes

Per sort, encara

falta molt perquè

la intel·ligència

artificial sigui

analítica i reflexiva

passen desapercebuts per als usuaris, però que incideixen en aquestes posicions discriminatòries.

Com ara...?

Doncs els *chatbots* com Alexia o Siri. Tots dos utilitzen veus de dona, i aquest fet perpetua el paper de servei que se li ha donat al gènere femení al llarg de la història. Aquest és un estereotip tan assumit socialment que el trobem absolutament normal. Això és el que hem de combatre.

Quin són els biaixos majoritaris que han detectat?

Els més nombrosos són els de gènere, però també és cert que és el biaix que més busquem a l'hora de provar les nostres solucions i també és el focus de diversos projectes nostres. Però en alguns casos, la discriminació de gènere va acompanyada d'altres tipus de discriminacions. Recordo un cas de biaix de gènere en què la discriminació augmentava si sumaves dona i raça. Es tractava d'un sistema de vigilància biomètrica que identificava els subjectes teòricament més sospitosos i, en aquest cas,

eren les dones de raça negra.

Han trobat algun altre tipus de biaix destacable?

Hem col·laborat amb Wikipedia en la detecció de forats culturals (*cultural gaps*), és a dir, zones geogràfiques amb menor cobertura que d'altres, i hem aportat eines per reduir-los amb la creació del Wikipedia Diversity Observatory.

La incorporació de la intel·ligència artificial als algoritmes computacionals d'aprenentatge permet, entre altres funcions, crear i alhora localitzar biaixos, però què passarà quan s'apliqui la intel·ligència artificial generativa? Encara es complicarà molt més?

La gran preocupació de la integració de la intel·ligència artificial és la privacitat de les dades que s'entreguen als algoritmes. I aquest és un tema extremadament important perquè hem de pensar que els sistemes reben cada vegada més dades per aprendre. El *big data* s'utilitza per entrenar models, i aquest models han de ser transparents i correctes per poder fer front als biaixos.



Eugenia Fuenmayor, en una de les sales de reunions de l'àrea digital del centre tecnològic Eurecat, de la qual és la directora científica, a Barcelona. ORIOL DURAN

Per tant, què s'ha de tenir en compte?

Els nous models de llenguatge són molt convincents i si la gent no és conscient que poden contenir errades o biaixos, ells mateixos faran de caixa de ressonància i divulgaran i impulsaran aquestes discriminacions al global de la societat. A hores d'ara, la Unió Europea està intentant fer una regulació sobre això, i a mitjans de juny va aprovar l'inici de negociacions per la que serà la primera llei del món que regula l'ús de la intel·ligència artificial, que es preveu que estigui llesta a finals d'aquest any. La prioritat d'aquesta normativa és garantir que els sistemes d'intel·ligència artificial que s'utilitzin a la UE siguin segurs, transparents, amb traçabilitat, no discriminatoris i respectuosos amb el medi ambient. A més, sembla que també vol obligar que aquests sistemes estiguin supervisats per persones i que aquesta tasca no es faci de manera automatitzada.

La nova llei diu alguna cosa sobre qui ha d'estar implicat en el disseny dels algorismes?

Allò que sabem, per ara, és que la normativa europea indica que els desenvolupadors tindran molta responsabilitat, molta més que les companyies, en el desenvolupament de la intel·ligència artificial. Tanmateix, el problema segueix sent el mateix que tenim ara, i és que en el disseny i desenvolupament de la intel·ligència artificial hi intervé una dona per cada cinc homes. L'increment de la diversitat és ara més necessari que mai.

EL PERFIL

Investigadora i formadora

Eugenia Fuenmayor es va graduar en Informàtica l'any 1980, a Veneçuela. La seva, explica, va ser una promoció "totalment atípica": "Les dones érem gairebé el 90% dels estudiants, però quan vaig passar a fer el doctorat, ja vaig ser l'única dona matriculada." Fa vint-i-dos anys va decidir traslladar-se a Barcelona. Actualment, a més de la seva posició a Eurecat, imparteix l'assignatura Introducció a la

Pensa que podria ser una opció fixar quotes en les etapes de disseny i desenvolupament de la intel·ligència artificial?

Instaurar quotes que discriminin positivament la dona en aquest camp és molt difícil, perquè continuament estan sorgint noves tecnologies, i les dones són minoritàries en les carreres STEM (l'acrònim en anglès que fa referència als estudis de ciències, tecnologia, enginyeria i matemàtica). El

programació, a la Universitat Pompeu Fabra, amb una majoria d'alumnes del gènere masculí. Tot i això, se la percep satisfeta quan explica que a Eurecat "un 48% de les més de 700 persones que hi treballen són dones", una proporció que decreix lleugerament en l'àrea digital del centre tecnològic, on estan involucrats un centenar d'investigadors, tot i que "anem avançant", assegura.

problema s'ha d'atacar abans, des de l'escola i la família. Les estadístiques ens indiquen que hi ha tant talent en nens com en nenes, que a l'Estat espanyol es matriculen a les universitats més dones que homes, però que tan sols un 30% de dones trien estudis STEM i, dins d'aquests, el percentatge baixa si parlem d'intel·ligència artificial i computació. S'ha de trencar aquest cercle i crear referents femenins per a les nenes que comencen ara l'escola, explicar que les primeres persones que van escriure un algorisme i un compilador per a programació van ser dues dones, tot i que els seus noms no apareixen mai. A aquest oblit també ajuda el fet que la gran majoria de professors universitaris a les carreres STEM són homes.

Cada vegada són més nombrosos els col·lectius que demanen que l'ètica entri en el desenvolupament tecnològic, també d'aquells que es postulen contraris a un excés creixement de la intel·ligència artificial. Es poden posar barreres en aquest camp? L'evolució tecnològica no es

pot aturar, i l'ètica no pot quedar-ne al marge. La intel·ligència artificial canviarà moltes coses, per bé i per mal, però no té un raonament al darrere i, per sort, encara falta molt perquè sigui analítica. Però un pas més serà la intel·ligència artificial reflexiva, i aquesta sí que podrà tenir el control del sistema. Per tant, la nostra meta ha de ser construir les millors màquines possibles, però també tan equànimes com sigui possible, amb els humans mantenint-ne el control.

Les màquines podran suplir les persones algun dia?

Mai! Una màquina no té sentiments. Pots fer que simulin visió o olfacte, però mai experimentar allò que les persones perceben de passió, risc, amor, compassió... No hem de tenir por dels avenços tecnològics, però sí que hem de controlar el mal ús de la tecnologia i estar vigilants i avançar-nos a possibles discriminacions, minimitzant-les al màxim. Personalment, soc molt positiva en aquest sentit.

I això vol dir?

Que crec en la humanitat.