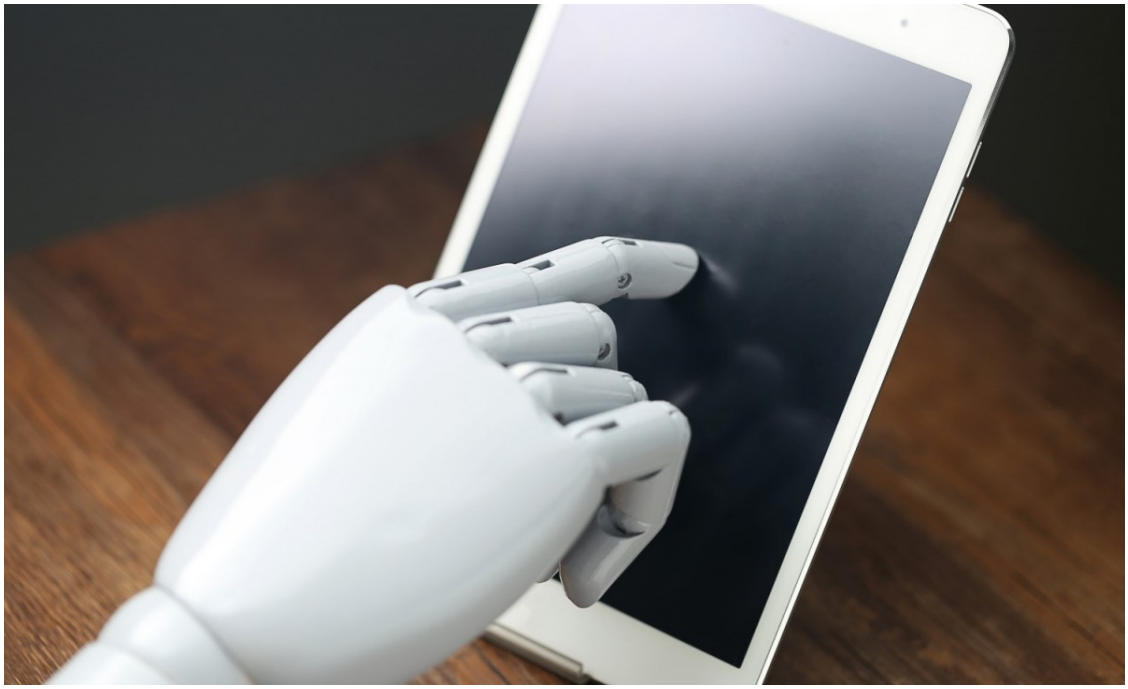
ECONOMIA | | Actualitzat el 17/09/2020 a les 12:00

L'ètica de la intel·ligència artificial: la tecnologia ens salvarà?

S'ha avançat molt en la inclusió de l'ètica en el disseny de la IA, però això no canvia els desequilibris de poder entre les empreses i els consumidors, entre empresaris i treballadors, o bé entre governs i ciutadans



Els sistemes basats en IA poden comportar certs riscos i tenir impactes negatius difícils d'anticipar, identificar o mesurar. | Adrià Costa

Una de les estratègies per controlar la propagació del SARS-CoV-2 és amb l'ús de tecnologies que alerten quan algú ha estat en contacte amb una persona infectada amb el virus. Algunes d'aquestes tecnologies prenen la forma d'aplicació de mòbil, Apps o aplicacions de rastreig de contactes. Actualment estan funcionant al voltant d'una cinquantena d'aplicacions globalment disponibles (https://www.top10vpn.com/research/investigations/covid-19-digital-rights-tracker/) . Tot i el seu potencial per ajudar a una detecció precoç dels casos, **aquestes aplicacions porten el risc implícit que les dades recollides siguin usades per uns altres fins diferents**.

Malgrat que la pandèmia de la CoViD-19 (https://www.pensem.cat/grans/temes/1335/repercussions/covid-19) s'espera que sigui transitòria, no hi ha cap certesa. Aquestes aplicacions recullen dades de caràcter personal i sensible, com ara els nostres moviments, les nostres interaccions socials i el nostre estat de salut, creant dubtes sobre com mantenir la privacitat dels usuaris i amb possibilitat de danys si aquestes dades cauen en males mans. **Desenvolupar aplicacions tecnològiques sense considerar les seves implicacions socials i ètiques pot comportar conseqüències costoses i perilloses**, desfavorint així els efectes beneficiosos a llarg termini.

Un grup d'investigadors en intel·ligència artificial (IA) a Suècia va desenvolupar un model informàtic (https://simassocc.org/) que simulava les conseqüències de les mesures polítiques per

frenar la corba del coronavirus. El que es va trobar va ser que **les aplicacions de rastreig de contactes són menys efectives del que es pretén**. "Utilitzar una aplicació de rastreig de contactes pot ser una eina que ajudi, però no és, com creuen ara moltes persones, una solució general que ens permeti aixecar totes les restriccions", digué Frank Dignum (https://www.politico.com/news/2020/04/24/coronavirus-contact-tracing-apps-206302) , professor de IA social de la Universitat d'Umeå (Suècia). En aquesta investigació es va comprovar que es requeriria que almenys un 60% de la població fes servir l'aplicació perquè fos efectiva. Així doncs, no es tracta tan sols d'un problema de privacitat, sinó també d'efectivitat.

Les aplicacions de rastreig són només un altre exemple d'una àrea en la qual es generalitza l'ús de la IA, com ja s'ha fet, per exemple, en vigilància predictiva, en treball social o en contractació de personal. **La IA té un atractiu que no tenen d'altres tecnologies**. Per què hi ha tanta lloança? Podríem començar per pensar a què ens referim quan parlem de "IA". Hi ha nombroses formes de definir-la, ja que és una denominació global. Una de les més acceptades és que la IA és una disciplina que estudia i desenvolupa artefactes computacionals que exhibeixen alguna faceta de comportament "intel·ligent". I hi ha tantes definicions d'"intel·ligència" com n'hi ha de IA. En aquest cas podríem dir que s'entén com a "intel·ligent" la capacitat d'un sistema per interpretar dades externes correctament, aprendre'n i fer servir els coneixements adquirits per assolir objectius específics i completar tasques a través d'una adaptació flexible. Així i tot, moltes de les aplicacions i sistemes que avui dia anomenem com a IA estan molt limitats en allò que poden aconseguir.

[noticiadiariambautor]93/122[/noticiadiariambautor]

En general, quan es parla de les capacitats de la IA, tendim a pensar en enginyers i executius de grans empreses tecnològiques, però la realitat és més complexa. Sovint s'ignoren els altres humans que hi ha al darrere, sobretot els que es dediquen a la recopilació, l'anotació i a la neteja de les dades, totes feines essencials que romanen invisibles. Aquest tipus de feina s'anomena "*ghost work*" ("treball fantasma"), terme encunyat per Mary L. Gray i Siddharth Suri al llibre (https://ghostwork.info/) que van publicar el 2019. **No sabem si aprenent d'unes dades** (dades que es puguin transformar en interpretables pels ordinadors), **les màquines basades en IA podran respondre correctament a entorns nous i impredictibles**.

I, el que és més important, quin és l'impacte que tindrà en les societats i en el medi ambient? **El potencial per automatitzar processos és alt, però això comportarà tota una sèrie de conseqüències -econòmiques, polítiques, socials, ambientals- imprevistes.** És més urgent que mai reflexionar en profunditat i analitzar quan és realment necessari crear un sistema completament nou enlloc de millorar-ne un de ja existent. A més, és important considerar quins són els criteris per decidir usar IA en comptes d'una tecnologia diferent i com fer-ho per incorporar-la en d'altres processos existents. Fàcilment trobem moltes empreses i individus amb idees brillants de sistemes de IA que ens ajudaran a fer la vida més fàcil, però quants d'aquests inclouen al seu pla de desenvolupament el manteniment profund del sistema: qui és el responsable quan el sistema cometi errors de biaix de gènere, d'ètnia, d'orientació sexual, etc., què fer quan el sistema s'escapa a qualsevol tipus de control social, què fer quan comet errors amb un origen complex de determinar? Podria ser que aquests sistemes de IA fossin una arma de doble tall?

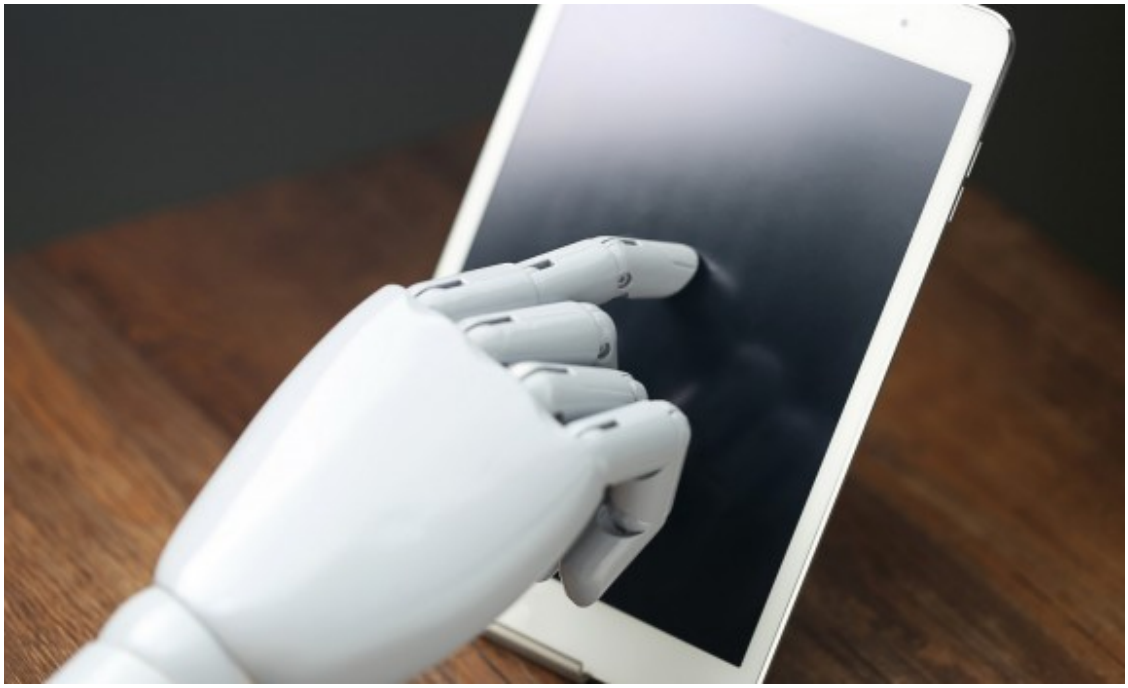
Precisament són totes aquestes preguntes les que han motivat que, en els últims quatre anys, haguem anat veient **diferents períodes -anomenats *onades*- de com es concebria una IA que fos "ètica"**. La primera onada se centrava en principis, guies i codis ètics sobre com haurien de ser les tecnologies de IA, i es va basar en allò que declaraven els filòsofs i especialistes en ètica. Aquests principis solen anar enfocats a orientacions "d'alt nivell", fet que els feia difícil d'aplicar a casos individuals i emfatitzaven l'ètica per sobre les regulacions. Normalment aquestes guies van implícites amb l'assumpció que la IA serà útil per resoldre problemes i van donar peu a la crítica que el moviment havia estat manipulat per grans empreses tecnològiques com a mitjà per evadir la intervenció reguladora.

La segona onada fou dirigida per informàtics teòrics, i es va orientar a trobar solucions tècniques per abordar qüestions d'equitat, biaixos i discriminació. Hi ha dos casos concrets que van contribuir al desenvolupament d'aquesta onada. Un d'ells fou un estudi (http://gendershades.org/) en el qual es descobrí que un *software* de reconeixement facial creat per Microsoft i IBM mostrava un major percentatge de falsos positius i falsos negatius en dones negres. L'altre estudi és una investigació de ProPublica (http://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing) en què es va analitzar una eina algorítmica de suport de decisions usada en jutjats dels Estats Units d'Amèrica, on es va trobar que els acusats negres tenien més propensió a ser incorrectament jutjats que els acusats blancs. Aquest focus en l'equitat tècnica és insuficient per entendre la complexitat de riscos i beneficis que ens presenta la IA, ja que no ens ajuda a pensar si, d'entrada, un sistema s'hauria de dissenyar o no.

La tercera onada va més enllà dels principis ètics i de solucions tècniques i se centra en mecanismes pràctics per rectificar els desequilibris de poder i aconseguir la justícia individual i social. Aquí és on es proposa la IA "justa", i es volen analitzar més profundament aplicacions i casos d'ús específics. En un any caracteritzat per les desigualtats socials en l'accés a la sanitat i les protestes del moviment de *Black Lives Matter*, aquesta onada es concentra en els problemes estructurals, inclosa la importància de "descolonitzar" (https://arxiv.org/abs/2007.04068) la IA. Es basa a canviar la visió més tècnica de la IA cap a una visió sociotècnica, és a dir, tenir en compte el context social en el qual les tecnologies es desenvolupen i s'implementen. D'aquesta forma, aquesta onada exterioritza d'altres problemes com, per exemple, l'impacte climàtic de la IA. Així i tot, aquest enfocament de la tercera onada no és àmpliament acceptat en tot el sector tecnològic.

[noticiadiariambautor]93/104[/noticiadiariambautor]

Tornant a les aplicacions de rastreig de contactes, hem de ser molt curosos i transparents sobre allò que estem fent i quines són les nostres pressuposicions. Tenir la capacitat de crear quelcom no porta implícit l'obligació de portar-ho a terme. La IA és una eina amb molt de potencial per ajudar-nos a mitigar alguns dels grans problemes de la humanitat, però **no ens podem afanyar a implementar-la a tot arreu sense abans haver-ne calibrat l'impacte**. Els sistemes basats en IA poden comportar certs riscos i tenir impactes negatius difícils d'anticipar, identificar o mesurar. Tot i haver avançat molt en la inclusió de l'ètica en el disseny, això no canvia els encara existents desequilibris de poder entre les empreses i els consumidors, els empresaris i els treballadors o entre els governs i els ciutadans. En aquest procés és important incorporar-hi d'altres perspectives, sobretot aquelles que han estat històricament silenciades, i també sospesar com haurien de ser aquests sistemes i si caldria d'entrada crear-ne de nous. L'expressió "sols el poble salva el poble" adquireix un nou sentit en aquest context: hem de repensar la il·lusió que la tecnologia ens salvarà.



Els sistemes basats en IA poden comportar certs riscos i tenir impactes negatius difícils d'anticipar, identificar o mesurar Foto: Freepik/xb100

Bibliografia:

Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. "*Machine bias* (https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing) ." ProPublica. 2016.

Bowles, Cennydd. *Future Ethics*. (https://nownext.studio/future-ethics) NowNext Press, 2018.

Buolamwini, Joy, and Timnit Gebru. "Gender shades: Intersectional accuracy disparities in commercial gender classification (https://www.media.mit.edu/publications/gender-shades-intersectional-accuracy-disparities-in-commercial-gender-classification/) ." In: *Conference on fairness, accountability and transparency*, pp. 77-91. 2018.

Coeckelbergh, Mark. *AI Ethics*. (https://mitpress.mit.edu/books/ai-ethics) MIT Press, 2020.

Dignum, Virginia. *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. (https://www.springer.com/gp/book/9783030303709) Springer Nature, 2019.

Gray, Mary L., and Siddharth Suri. *Ghost work: how to stop Silicon Valley from building a new global underclass*. (https://ghostwork.info/) Eamon Dolan Books, 2019.

Pasquale, Frank. *The Black Box Society: The Secret Algorithms That Control Money and Information*. (https://www.hup.harvard.edu/catalog.php?isbn=9780674970847) Harvard University Press, 2015.

Russell, Stuart, and Peter Norvig. *Artificial intelligence: a modern approach* (https://www.pearson.com/us/higher-education/program/Russell-Artificial-Intelligence-A-Modern-Approach-4th-Edition/PGM1263338.html) . Hoboken: Pearson, 2020.

Whittaker, Meredith, Kate Crawford, Roel Dobbe, Genevieve Fried, Elizabeth Kaziunas, Varoon Mathur, Sarah Mysers West, Rashida Richardson, Jason Schultz, and Oscar Schwartz. *AI Now report 2018* (https://ainowinstitute.org/AI_Now_2018_Report.pdf) . New York: AI Now Institute at New York University, 2018.