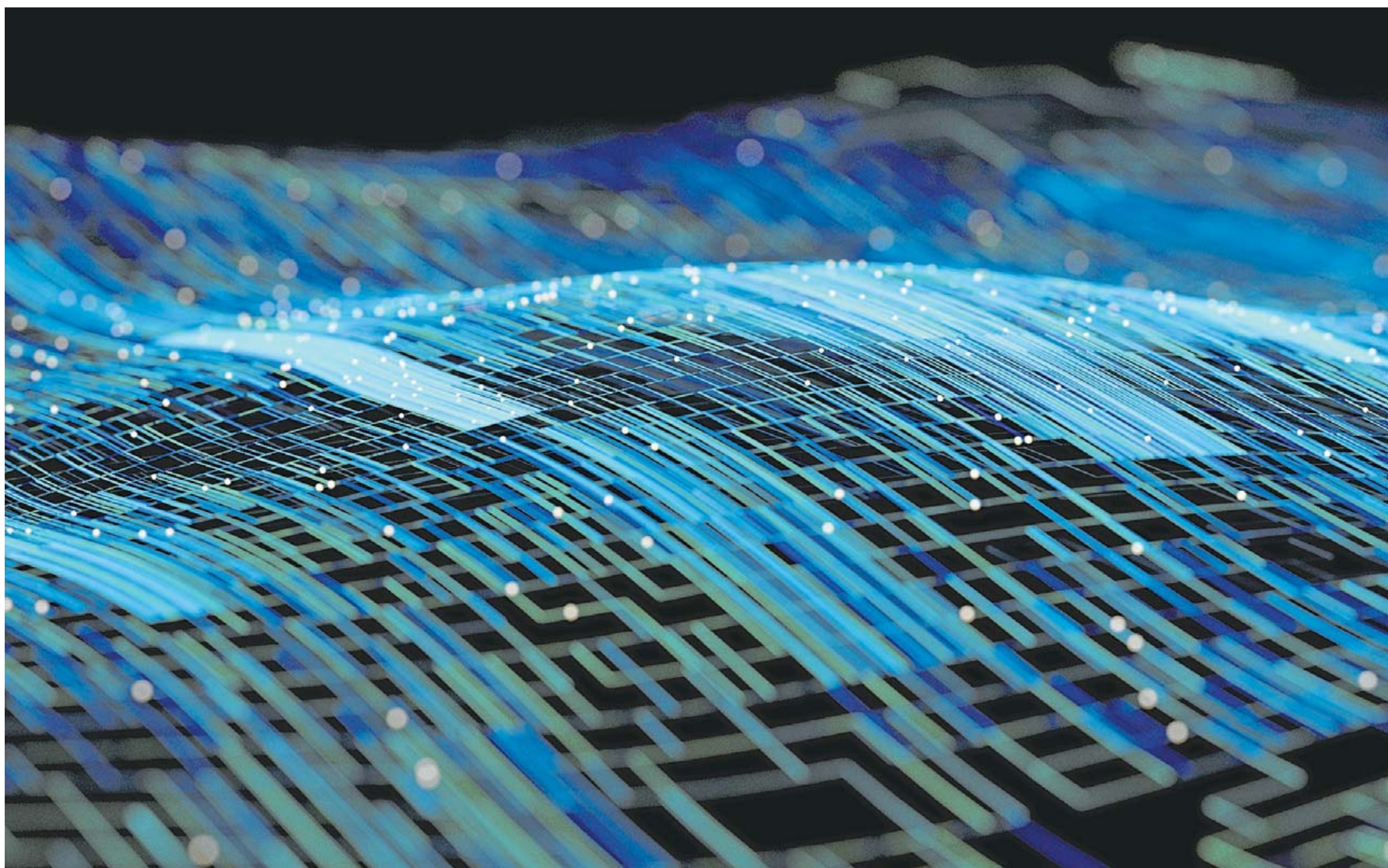
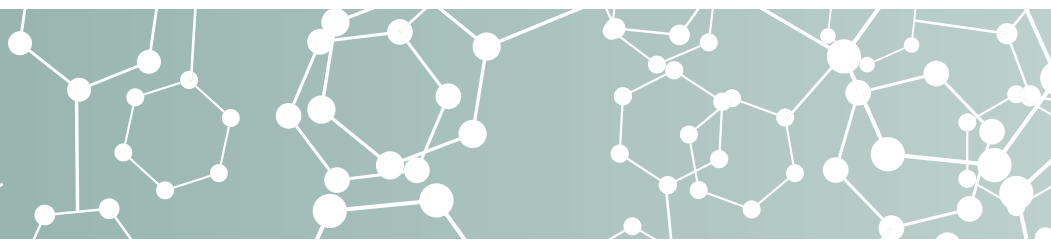


ara ciència



Es pot crear una intel·ligència artificial ètica i responsable?

Hi ha el risc que la tecnologia perpetui biaixos en la recerca, els judicis o la selecció de personal

Cristina Sáez

Tots i cada un dels éssers vius que habitem la Terra som únics i irrepetibles, almenys biològicament parlant. La nostra singularitat està codificada en un còctel de milions de combinacions de molècules al nostre ADN. La variació és intrínseca a la vida i ha estat el motor de l'evolució. Tanmateix, si bé algunes diferències es tradueixen en atributs o característiques diferents, com ara el color dels cabells, d'altres poden predisposar-nos a patir malalties.

El problema és com discernir les unes de les altres, les que són beneficioses o simplement innòcues d'aquelles que poden ser perjudicials. Tot i que les tecnologies i tècniques desenvolupades els dar-

rers anys per llegir el nostre codi genètic han permès als científics identificar cada cop amb més precisió aquestes variants, poder seqüenciar el genoma, llegir-lo, no implica entendre què volen dir les mutacions, ni tampoc ser capaços d'interpretar-les per poder assenyalar les que són realment importants.

D'aquí la importància de l'eina dissenyada per dos investigadors del Centre de Regulació Genòmica (CRG) de Barcelona, Mafalda Dias i Jonathan Frazer. L'algoritme que han desenvolupat, anomenat EVE, permet escodrinyar patrons de variació genètica en centenars de milers d'espècies animals no humanes per poder aprendre quines sí que són rellevants i fer prediccions acurades per als pacients sobre les mutacions que tenen. I això és cru-

cial no tan sols per ser molt més precisos en el diagnòstic, sinó també per poder fer pronòstics més encertats i triar el tractament més adient per a cada individu.

La tecnologia IA que han generat, de xarxes neuronals basades en un tipus sofisticat d'aprenentatge màquina, ja està implementada en hospitals dels Estats Units. És un exemple, de fet, de com l'aplicació de tecnologies d'intel·ligència artificial pot fer contribucions cabdals en ciència i recerca, i propiciar avenços en el coneixement, com ja van posar de manifest els premis Nobel d'enguany en les categories de química i de física, que van premiar, precisament, aquesta potentíssima eina i avenços realitzats amb ella.

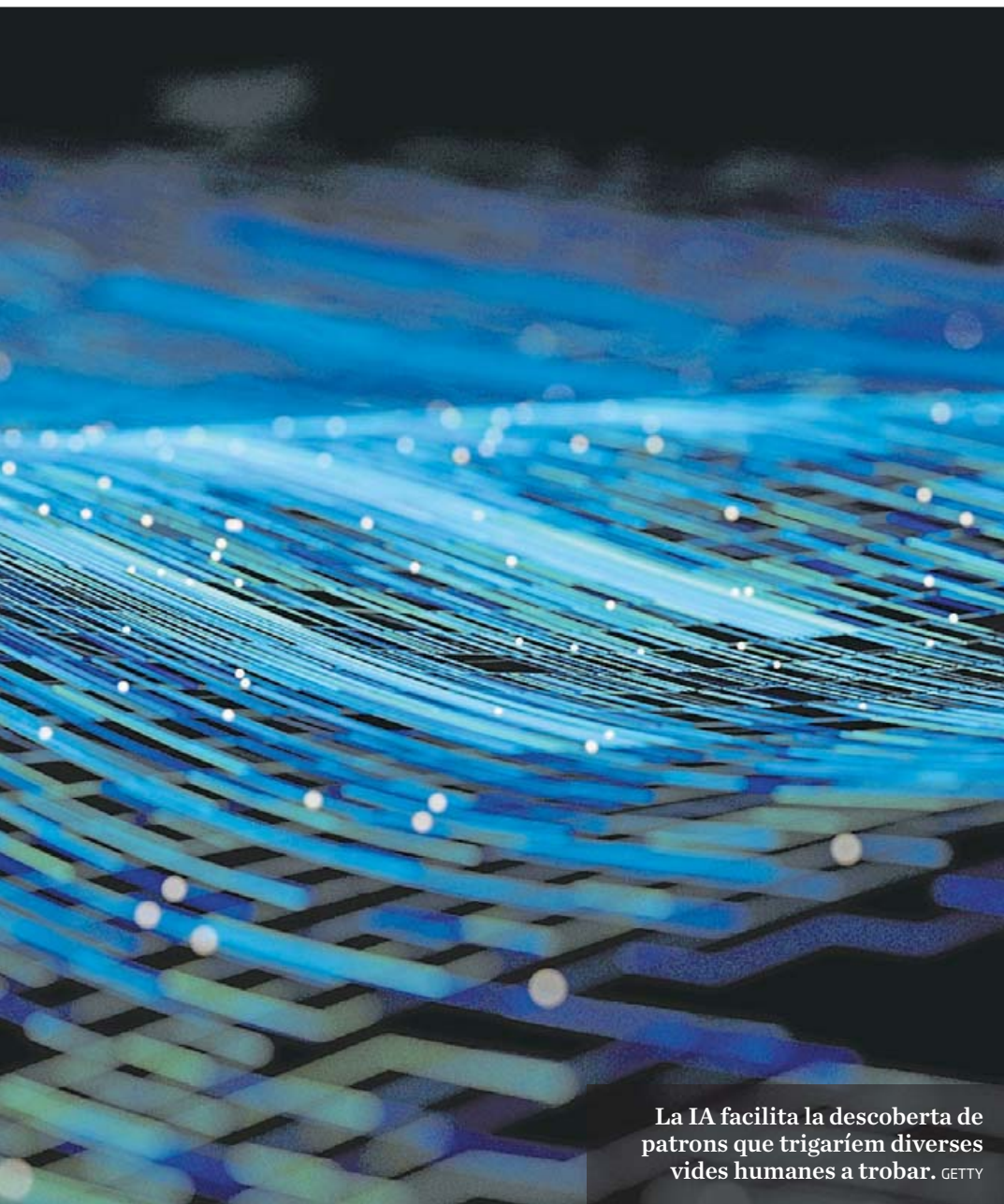
Fer servir intel·ligència artificial permet automatitzar tasques repetitives;

analitzar grans volums de dades amb una velocitat i una precisió inigualables, i facilitar la descoberta de patrons i correlacions que trigàrem diverses vides humanes a trobar. D'exemples n'hi ha molts en camps com la biomedicina, l'astronomia, la química, la física i les ciències socials. Totes aquestes potencialitats han saltat al domini públic.

Una eina amb clarobscur

"Hi ha un abans i un després en la IA arran de ChatGPT", diu l'investigador Icrea Alfonso Valencia, director del departament de ciències de la vida del Barcelona Supercomputing Center (BSC) i director de l'Institut Nacional Espanyol de Bioinformàtica (INB/Elixir-es). "Avui dia vivim en un pic entre les ex-

TECNOLOGIA



La IA facilita la descoberta de patrons que trigariem diverses vides humanes a trobar. GETTY

pectatives i les notícies sobre aquesta tecnologia, que sembla que serà a tot arreu i ubiqua a tots els àmbits de la ciència i la recerca”, diu aquest expert, que, tanmateix, precisa: “Vindran temps menys optimistes i ens adonarem de les limitacions que té, que no sabem si són controlables o no”.

L'ús de la IA, apunta Valencia, també presenta reptes i clarobscurs, com ara qüestions ètiques i de privacitat relacionades amb l'ús de les dades, així com la necessitat de garantir la transparència i el funcionament dels algoritmes. “La IA és una eina molt potent per aconseguir avenços socials en molts àmbits”, diu l'enginyera informàtica Àtia Cortés, del departament de ciències de la vida del BSC i membre del Comitè de Bioètica d'Espanya, que treballa en projectes de vehicles autònoms i robots assistencials. “Pot ajudar a potenciar les capacitats humanes i donar suport a una millor presa de decisions”, afirma.

Ara bé, puntualitza, l'afany humà per avançar, per aconseguir noves fites i fer-ho ràpid, sumat a l'enorme competitivitat entre empreses que treballen amb IA, sovint comporta “renúncies” o efectes secundaris indesitjables. Per començar, pel que fa a la privadesa. Cortés en posa un exemple: “Vull dissenyar cotxes autònoms amb molts sensors perquè així faig un sistema més robust que pot prendre decisions amb més rapidesa i eficàcia”.

D'altra banda, “això propicia que es recol·lecti un munt d'informació privada de les persones a dintre del cotxe que pot acabar sent un perjudici en temes de privacitat, per exemple de cara a ciberatacs”. A més, afegeix: “En les infraestructures públiques que també capten aquestes dades, què passa? Qui se les queda, qui les reutilitza i amb quina finalitat?”

Les dades són el nou or. Perquè els algoritmes estan entrenats amb elles i de la seva qualitat depèn la dels resultats de

la intel·ligència artificial. “Els algoritmes fan una representació d'un model del món a partir de dades que són esbiaixades”, apunta l'enginyer informàtic Josep Maria Ganyet, professor de la Universitat Pompeu Fabra expert en IA. “Si no corregeixes els biaixos, el que generis estarà esbiaixat”, diu. Cada cop que entrem en una xarxa social, en funció de si som home o dona i de la nostra edat i del que l'algoritme ha recopilat a partir dels nostres *likes*, ens mostrarà allò que suposadament ens interessa. A les dones, cuina, criança i bellesa; als homes, culte al cos, esports de risc, economia, vehicles. El mateix passa a les plataformes digitals audiovisuals.

Tot i que en el cas d'aquests continguts l'ús de dades de la IA pot semblar inofensiu o anecdòtic, en àmbits com ara el de diagnòstic i tractaments mèdics, de selecció de personal o en judicis, on es fan servir eines amb IA, pot no ser-ho. “Si no es té cura que la mostra estigui equilibrada, sense cap mena de discriminació de cap grup social, i que tinguin en compte les bretxes que hi ha en el món digital, on tot-hom no està representat igual, aquestes eines poden arribar a ser perilloses”, diu Ariadna Font, cofundadora i directora de l'*start-up* Alinia.AI i creadora de l'àrea ètica algorítmica de Twitter abans que passés a ser X. Recorda el cas d'una campanya de contractació de personal d'Amazon feta amb un sistema de reclutament basat en IA que “tenia molts biaixos contra les dones per a rols tècnics”: “Van haver de deixar de fer servir aquest sistema”, explica.

En recerca en el camp de la salut, recorda Cortés, tot i els esforços per corregir-los, encara hi ha biaixos que alimenten les eines d'IA. Per començar, perquè fins fa poques dècades les dones estaven excloses dels assajos clínics, i ja des de les fases preclíniques només es feien servir models animals masculins. Això va fer que tant els fàrmacs desenvolupats –amb efectes secundaris indesitjats o no eficaces en dones– com la detecció de símptomes de malalties no fos acurada. “Aquests biaixos en recerca afecten sempre els col·lectius més minoritaris: ètnics, demo-

gràfics, per gènere o per estatus socioeconòmic”, coincideixen els experts.

Més transparència

Un altre aspecte espinós de la IA és el de la responsabilitat i la transparència de l'algoritme. Tornant a l'àrea mèdica, quan es desenvolupen sistemes basats en intel·ligència artificial per diagnosticar una malaltia amb una elevada precisió, entra en qüestionament el judici del professional. “Si la IA em diu que hi ha un 88% de probabilitats que sigui un tumor, però com a metge en tinc dubtes, què he de fer, no donar-li tractament i que temps després es demostrï que sí que era un càncer? O donar-n'hi per precaució, tot i els efectes secundaris perjudicials si no pateix cap tumor?”, planteja Cortés posant-se en la pell d'un professional de la salut.

Per a aquesta experta en ètica, el problema de fons és la no transparència dels algoritmes, que impedeix entendre'n el

disseny, com han estat alimentats i com funcionen. “Per això –diu– és fonamental que aquestes eines portin documentació comprensible per a tota mena de públic, que expliqui el procés que s'ha seguit per arribar a aquells resultats i

també les limitacions del sistema. Només així realment poden ser una eina de suport de les decisions i permetre que els metges prenguin eleccions de la manera més informada possible”.

Font ressalta que les decisions no les prenen les màquines, sinó que ho han de fer les persones, cosa que no ha passat en casos recents, com va evidenciar el cas de les prestacions socials de persones en situació de vulnerabilitat extrema a Holanda, considerades erròniament com a frauds per un algoritme del govern. I la història es repeteix en jutjats, bancs i hospitals quan qui ha de prendre decisions relega la responsabilitat a les màquines. “La IA no ha de prendre decisions per les persones, sinó donar dades i eines per adoptar decisions més ràpides i de més qualitat”, corregeix Font, que explica que per a “un metge que ha de mirar no sé quantes radiografies el que pot fer la IA és filtrar-les i donar-l'hi les quatre que poden ser més problemàtiques perquè primer se centri en elles”. “La responsabilitat última sempre és de la persona”, conclou.

Ganyet agafa aquest argument i va més enllà: “En alguns àmbits la presa de decisions pot ser problemàtica, com ara en justícia, i les decisions han de ser explicables per llei. Amb la IA no pots argumentar res, has de fer un acte de fe”. Per això, considera: “Cal plantejar-se fins on estem disposats a supeditar la nostra decisió a la màquina”. A més, afegeix que ja hi ha científics treballant en l'explicabilitat de la IA generativa.

Al final, coincideixen tots els experts, cal crear algoritmes responsables. “Hem d'apostar per una IA que es caracteritzi per la transparència, l'explicabilitat, la rendició de comptes, la privadesa de les dades, i que la demanem. Són els principis que caldria aplicar a aquestes eines per alinear-les amb els valors humans. Els sistemes d'IA no tenen ètica ni moral si no és que tu els hi marques”.

La IA no pot prendre decisions, sinó ajudar els humans a fer-ho més ràpid i amb més qualitat

Tecnologies a debat

Per reflexionar sobre els desafiaments de la intel·ligència artificial aplicada a l'àmbit de la recerca, la Comissió de Tecnologia i Recerca de la Coordinadora Catalana de Fundacions (CCF) –una entitat que representa prop de 800 fundacions catalanes d'àmbits tan diversos com la salut, l'educació i la recerca– va organitzar recentment unes jornades sobre l'impacte transformador i els reptes ètics d'aquestes tecnologies. L'objectiu, en paraules de la presidenta de la coordinadora, Eugènia Bieto, era “generar coneixement i debat”, així com fomentar el contacte i la xarxa entre experts de diversos àmbits. A l'acte hi van assistir un centenar de persones, la majoria caps i coordinadors de projectes de recerca de l'ecosistema recercaire català, directores de fundacions, responsables d'innovació i desenvolupament, i analistes de dades. Des de la comissió de tecnologia i recerca de la CCF es va plantejar organitzar una jornada sobre l'impacte de la IA en la recerca, atès que “es tracta d'una tecnologia que ha tingut un impacte immediat, profund i transformador en el dia a dia dels investigadors”. Amb aquesta trobada, des de la CCF es volia “donar resposta a qüestions clau [...], com ara de què és capaç la IA, com pot ajudar a l'hora de dur a terme un projecte de recerca”, però també “quines limitacions i interrogants planteja”.