Les idees: Humà i Intel·ligència Artificial

Queralt Prat i Pubill

No hi ha res més pràctic que una bona teoria- Kurt Lewin

Res és més poderós que una idea a la que li ha arribat la seva hora. - Victor Hugo

Tard o d'hora, són les idees, no els interessos creats, les que són perilloses per al bé o per al mal.
-John Maynard Keynes, *The General Theory of Employment, Interest and Money*, ch. 24, p. 383 (1935)

Escric aquest text en un moment d'avanç accelerat de la Intel·ligència artificial ("IA"). Aquest fenomen no només afecta a la nostra manera de treballar o de comunicar-nos, sinó que incideix en la nostra concepció de coneixement, la realitat i el món. En resum, les nostres concepcions i per tant afecta a la creació del nostre futur. Dos aspectes fonamentals m'animen, un, la comprensió, de quines son les idees que es socialitzen sobre la IA i la segona, algunes reflexions sobre la nostra concepció de què vol dir ser humà quan els sistemes de IA poden desenvolupar tasques "intel·ligents" millor que nosaltres.

Recordant les paraules de JM Keynes i Victor Hugo expliciten que les idees tenen un poder més gran del que ens imaginem. També que les teories son pràctiques perquè ens permeten donar sentit al món de manera pràctica. L'entrada de la IA en la nostra vida ens està demanant repensar què vol dir ser humà, quina humanitat volem construir i per tant com ens hauríem d'orientar com a espècie. Així, aquestes reflexions no són un exercici teòric, allunyat de la vida diària que estem vivint, sinó que incideixen a nivell pràctic de com el coneixement, la ciència, la tecnologia es desenvolupen amb IA i com això té un impacte en com pensem i organitzem les nostres vides i per tant en l'orientació dels sistemes de valors col·lectius. Aquests dos elements, els valors i el coneixement orienten la nostra prosperitat i més brutalment, la nostra supervivència, per tant, entendre com pensem sobre aquesta tecnologia: la IA, és crucial.

Fa més de 25 anys, que a nivell pràctic, em vaig interessar per entendre què vol dir conèixer el món, què és la realitat, és a dir, l'epistemologia. En aquells moments, tot era molt més simple, m'interessaven els textos de saviesa, m'explicaven una manera diferent d'entendre el món. A l'inici, de manera fortuita investigant en una llibreria a Londres amb l'Advaita, després amb el Budisme Zen i posteriorment, cinc anys més tard vaig començar a ampliar, gràcies a la guia i el treball metòdic del CETR, el meu coneixement de totes les altres tradicions, com per exemple, la cristiana, la musulmana i la taoista.

L'extraordinària aportació del CETR ha estat explicitar com separar les configuracions culturals, els sistemes de valors, d'aquestes tradicions, de la saviesa que volen comunicar. Per fer això, a l'inici, vam necessitar fer un treball epistemològic. Per tal de poder aprofundir en aquest treball de comprensió personal de la saviesa així com la capacitat per comunicar-ho a altres. Poder entendre i desenvolupar la Qualitat Humana, això que ens intenten mostrar les tradicions, aquesta capacitat humana de viure en el món més enllà de les comprensions automàtiques provinents de la estructura psicològica, resultat de tota una sèrie d'atzars que em situen en un espai-temps determinat, dins d'una cultura, dins d'una família amb una altre sèrie d'atzars personals, experiències, memòries, expectatives, desitjos i pors ha estat un gran regal.

Així, per poder fer aquesta investigació dels textos de saviesa s'ha de qüestionar les hipòtesi de base des d'on es fa la lectura, sinó treballarem sotmesos a la nostra configuració cultural, una presó epistemològica, de la que no es pot escapar, a no ser què un la conegui. Una de les creacions fonamentals del Marià Corbí (1983), ha estat establir com a hipòtesi del seu treball d'investigació que l'antropologia de l'animal humà ha d'estar basada en fets científics no en especulacions. Així, començar el desenvolupament teòric definint a l'animal humà com constituït pel llenguatge, és a dir, modelat en la seva comprensió del món i en la seva actuació depenent de les necessitats de supervivència del col·lectiu, significa que l'animal humà és flexible en el que valora i per tant en com actua. Flexibilitat no vol dir què tot val, què tot és relatiu. Vol dir que l'animal humà com a col·lectiu crearà aquells valors que li asseguren la supervivència, de manera més o menys conscient. Si no és capaç, llavors, aquella civilització, o fins hi tot l'espècie no prosperarà.

Aquesta flexibilitat en la valoració i per tant, en l'actuació, és possible perquè els humans tenim un sistema de comunicació on el significat de les coses del món es col·loca en el significant (els sons de les paraules) i es separa directament del seu significat, perquè es crea una distància objectiva que fa que la cosa en sí pugui tenir molts significats (de Saussure, 1959). Aquest tipus de llenguatge és molt especial, cap altre animal terrestre te aquesta estructura. La resta d'animals poden tenir llenguatges, però cap està constituït en el format humà com una triada que es representa en una relació humà-paraula-món. Això és el que ens atorga la nostra flexibilitat i la raó per la qual hem desenvolupat, entre d'altres, ciència, art i religions.

Ha estat impressionant copsar com aquesta decisió metodològica, de definir l'antropologia seguint les troballes reconegudes a nivell antropològic i lingüístic, continua essent tant obviada en el món acadèmic i social com fa més de 40 anys quan Corbí ja la va postular. Aquest principi antropològic, que defineix a l'humà com a constituït pel llenguatge, encara que de manera minoritària present a les ciències socials, principalment a la sociologia del coneixement, historia del coneixement i estudis tecnocientífics, no ha arribat a tenir impacte suficient per esdevenir rectora de noves propostes que arribin al nucli de la filosofia, ni de l'ètica i tampoc, és clar, a la religió. Evidentment, al no arribar al moll de les disciplines, no s'han pogut desenvolupar les conseqüències socials, polítiques i econòmiques d'aquesta conceptualització, encara que hi sigui present a nivell teòric. Encara a nivell social, cultural i fins hi tot en la majoria de les teoritzacions de les ciències socials es pensa o directament, no es qüestiona la hipòtesi primera, de l'animal humà com un compost de cos i raó o cos i esperit. En el món de la IA, ja ho explicitarem més endavant, es conceptualitza l'ésser humà com intel·ligència, diguem que és una destil·lació del concepte "raó" i "racionalitat", com un coneixement deslligat de l'experiència humana, diríem com un oracle o com un Déu, com si el coneixement estigués definit en un àmbit i s'hi pogués accedir amb la intel·ligència. Utilitzar hipòtesis com cos i raó o cos i esperit és construir la ciència des de la metafísica. Si aquestes hipòtesis estan en la base de totes les nostres creacions teòriques, llavors totes les nostres creacions científiques son clarament vulnerables. Això és un problema greu.

Considerar a l'animal humà com constituït pel llenguatge seria una revolució copernicana en les ciències socials, però és clar, llavors disciplines, per exemple, com l'ètica, serien rellevades a disquisicions de diletants, és a dir no rellevants, perquè es faria palès que formulacions racionals de valor, el que actualment és l'ètica en la nostra societat, son tan efectives com pregar als déus de l'Olimp[1]. Deixaríem d'invertir energies i capital intel·lectual en disciplines com la "ètica dels negocis" o "l'ètica de la IA" i ens enfocaríem a entendre com crear motivacions de valor que

[1] Si l'ésser humà és un animal racional, llavors principis d'actuació racionals poden guiar a l'humà. Però si conceptualitzem que l'ésser humà és un animal, llavors l'eix de l'actuació animal, és el sentir, com tots els altres animals. Evidentment, en el cas de l'humà, és un sentir que te lògica, causa-efecte, incertesa, etc. per tant, per orientar a l'animal humà no és suficient mitjançant la racionalitat quan està totalment subordinat al sentir.

fossin efectives en les organitzacions, a nivell social i polític i com d'aquí es podria derivar una ètica[2].

Comprendre a l'ésser humà com constituït pel llenguatge ja transpúa una flexibilitat immensa. Si ho comparem a altres espècies animals, veiem que és una característica única. Una flexibilitat extraordinària i la font del nostre èxit com superdepredadors a la Terra. Estar constituït pel llenguatge vol dir que aquesta constitució, el que som, el que l'ésser humà considera adequat com actuació, té la capacitat de canviar infinitament. Mai com una decisió arbitraria, sinó sempre subjecte a les formes de viure. Les diferents constitucions humanes es reflecteixen en les valoracions dels grups humans i en els significats de les paraules. Aquesta flexibilitat d'adaptació i creació de noves possibilitats no existeix en cap altre animal terrestre. És fonamental i delicada. Si els humans no estan constituïts adequadament no seran capaços de sobreviure en l'entorn. Això ja passa. En la nostra espècie conviuen simultàniament moltes configuracions humanes.

Per tant, aquesta constitució, configuració de les valoracions humanes, es defineix a nivell col·lectiu i no és resultat d'un procés participatiu, ni d'una enquesta, ni d'un mercat, ni de la decisió d'un primer ministre, un dictador o un rei. Sinó el resultat de molts atzars, de contribucions disparatades d'humans, uns més poderosos que altres, de processos llargs més enllà d'una vida humana, i sobretot, fins ara, d'una manca de coneixement que tot això estava passant. M'atreviria a dir que tenim aquesta manca de coneixement perquè tenim aquestes creences del què és l'humà basades en teories metafísiques, com si vingués del cel, som un "esperit" som una "raó", som una "intel·ligència". Si pensem que les configuracions valorals ens venen donades, o revelades per tal com és la natura de les coses, no hi ha espai per imaginar nous tipus de valoracions ni la necessitat de crear-les.

Tots els animals estan plenament configurats per sobreviure gràcies a la informació genètica que hereten, nosaltres no. Els humans tenim una "deficiència" que és un tresor. Nosaltres no tenim configurat com hem d'actuar per sobreviure genèticament, ni com ens hem d'organitzar, ni com ens hem de relacionar entre nosaltres (Corbí, 1983). La teorització del Marià Corbí, que no està materialitzada en les seves conseqüències a nivell social, ens permet ésser conscients que aquestes constitucions humanes han de canviar i canvien i que les configuracions de valors han de ser adequades als eixos amb els quals un col·lectiu sobreviu si no, no es pot prosperar. Per tant, **la prosperitat no és una fita tecnològica o cultural**, si no una fita que te a veure amb **la capacitat de constituir uns humans** per ésser capaços de prosperar en unes condicions adequades, és a dir, **és la conjunció axiològica** i de manera de sobreviure (tecnològica i cultural) que permet a les comunitats prosperar. Així, les nostres valoracions col·lectives, com el que definim com a intel·ligència no són atemporals, si no que son dependents que som un animal humà i de com sobrevivim en l'entorn.

Aquest element axiològic, aparentment irrellevant en el nostre món científic i cultural, és en realitat una força poderosa i indomable que condiciona la capacitat de comunitats i individus per prosperar. Dedicar esforços i recursos a entendre aquestes configuracions axiològiques que constitueixen als humans és tan important com la frontera tecnològica de la física quàntica, la biotecnologia o la Intel·ligència Artificial, per dir alguns camps científics. Encara m'atreviria a dir que **és més important que les nostres creacions científiques, perquè les configuracions axiològiques estructuren com les nostres creacions són utilitzades, i tenint en compte que la nostra ciència i la tecnologia és cada vegada més poderosa, això vol dir que hem de ser capaços de construir configuracions axiològiques que beneficiïn la vida no que portin a l'extinció de la humanitat.**

[2] Avui dia l'ètica es presenta com una formulació racional deslligada d'un sentir d'un projecte humà compartit.

Des d'altres disciplines han arribat a conclusions similars. Disciplines com els estudis de la filosofia de la ciència, la historia de la ciència i la sociologia del coneixement s'ha interessat en entendre com els humans construïm el coneixement, com és diferent de les creences i com està relacionat amb les comunitats d'humans que han creat aquell coneixement. En els anys 70, Thomas Kuhn (1970) va començar a sacsejar els pilars de la construcció del coneixement científic com a descripció del món. Va ser Boyle al 1660 qui va definir els eixos racionals que han desenvolupat el coneixement científic tal com el coneixem avui. Segons Boyle, per poder entendre el món, per poder descriure'l fefaentment hem de seguir un mètode: 1. fent experiments - tecnologia material 2. necessari que hi haguessin testimonis- tecnologia literària 3. testimonis fossin independents - tecnologia social (Shapin, 1984). En el segle XX la sociologia de la ciència ha conclòs que el coneixement científic no reflecteix la natura, tal com postulava Boyle, sinó que és una eina pràctica per "fer comprensible" el món (Bloor, 1991). Així, no és una descripció, més aviat una modelació, si el coneixement funciona a la pràctica, llavors diríem que allò és "veritat", sinó és "fals". El que es "descobreix", la tecnologia que es desenvolupa no és la "millor" tecnologia, com si hi hagués un camí predeterminat de desenvolupament sinó que es depenent de factors històrics (Callon, 2010) i per tant, la ciència i la innovació no té una dinàmica pròpia sinó que està totalment condicionada per la situació social. La visió de la ciència com autònoma i deslligada de la societat té implicacions polítiques perquè permet mantenir i perpetuar el status quo dominant (Law, 2017).

Des d'una aproximació feminista, Haraway (1988) ha demostrat empíricament com els mètodes i procediments per fer la feina científica portaven altres agendes, valors, coneixements, etc que afectaven el que es creava i com es creava, és a dir res era tan asèptic, neutra i fora de disputa tal com Boyle postulava. Tot el coneixement i tots els mètodes estant "situats", reflecteixen el lloc i reprodueixen les agendes socials. En la nostra societat pensem que la ciència és "objectiva", "la veritat", però Haraway (1988) argumenta que "aquesta mirada des d'enlloc" és com un "truc de Déu" que pot veure tot des d'enlloc, imparcial, sense tenir en compte tots els elements humans qui hi influeixen en el seu desenvolupament. Els termes "objectivitat" i "subjectivitat" perden les seves arrels, encara que Haraway manté la paraula objectivitat sempre tenint en compte dos aspectes (1) reconèixer que el quefer científic està situat (2) analitzar críticament aquest fet "situat". Per tant, per Haraway (1988) s'ha de descartar la ficció que la ciència sigui "neutra" i per tant, que és una descripció de la realitat, però no acaba de trencar el concepte d'objectivitat, on invariablement s' assumeix que una descripció fidedigna és possible, encara que ella no ho cregui. I és que el trencament de la possibilitat d'una descripció acreditada "objectiva" pot portar a conseqüències no desitjades. Si no hi ha res "creïble", què queda? hi ha molt filòsofs que s'han dedicat a reflexionar sobre l'impacte d'aquesta recerca i la incidència a nivell polític i en els valors.

Crítics d'aquesta postura veuen un atac a la racionalitat i per tant, segons ells, un retorn a la irracionalitat. Però, la repercussió d'aquestes investigacions porten a la conceptualització de l'ésser humà com un actor material-semiòtic, on el material i el que dona i crea sentit estan íntimament unit i que és gràcies a la interacció social, mitjançant la parla, que les fronteres del que és, son creades. Per tant, la ciència és una tecnologia semiòtica. I el que anomenem objectivitat és només una al·legoria de la ideologia que governa (Haraway, 1988). La mirada "objectiva", per tant, és la subjugada a l'ordre imperant no una "descripció" "autèntica" de la realitat. Aquesta disciplina, resultat de les investigacions empíriques sobre la ciència ha desenvolupat tot un eix teòric on es fa palès a nivell pràctic una epistemologia no mítica i una concepció del ésser humà fora de les coordenades comunes socials de cos i ànima o cos i esperit. Podríem suggerir que l'estudi del treball científic, des de la mirada feminista i l'epistemologia del coneixement, els ha portat a qüestionar la mateixa realitat de "l'existència" de l'individu, de la natura i de la societat. La teoria de les xarxes d'actors ("*Actor Network Theory*" ANT en anglès)

defineix aquesta "existència" semiòtica i per tant, modelada, i per tant, compresa transforma la nostra realitat del món (Latour, 2005).

Aquestes reflexions continuen encara, Law (2017) argumenta que la relació entre la societat i la ciència és bidireccional i estudiar aquesta situació és important. Invariablement aquestes aproximacions sacsejant la hipòtesi de descripció veraç de la ciència ha creat una forta oposició i investigadors com Harding (2017) han dedicat extenses reflexions a explicitar i argumentar perquè no es pot parlar d'objectivitat ni de subjectivitat per entendre com el humà entén el món. La física no és una opinió però al mateix temps tampoc és absoluta, una veritat que descriu el món[3]. Ha estat difícil conceptualitzar d'una manera rigorosa i conceptualment clara com el coneixement és una modelització del món que funciona, per tant, estrictament no és ni objectiva ni subjectiva perquè no tenim aquesta possibilitat. L'objectivitat assumeix que la nostra comprensió de la realitat és una descripció de la realitat i això no és possible.

Marià Corbí, ha aprofundit els seus estudis en entendre els valors. El món dels valors també és modelació humana, no hi ha un món de valors definit, veraç, autèntic, cert. Les configuracions axiològiques que han funcionat són els sistemes de valors que han permès a l'espècie progressar. Simplificant el desenvolupament teòric del Marià Corbí, podríem dir que fins ara hem viscut tres tipus de configuracions axiològiques que han organitzat la simbiosis dels humans: el mite, les religions i les ideologies. L'ordre cronològic que he descrit depèn de quan van ser efectives de manera dominant a nivell social. Avui dia, només tribus perdudes caçadores recol·lectores poden tenir mites. Les religions, nascudes de la revolució agrícola, encara son efectives en alguns països no occidentals encara com a estructuradores inqüestionables de la realitat, i les ideologies, bé potser només a Corea del Nord és possible que es mantinguin com a vàlides. En la resta del món, estem desestructurats axiològicament, hi ha una barreja de religions, ideologies i pensament alternatiu, tot barrejat, resultat de la globalització. Hi ha molt pocs llocs al món suficientment aïllats i amb formes pre-industrials de supervivència que no tinguin aquest "menú" axiològic, és a dir, aquesta barreja de possibilitats per ajudar al col·lectiu a prosperar.

Avui en dia, la prosperitat del col·lectiu humà no prové de la capacitat de caçar i recol·lectar fruits, o de la fertilitat de les terres i accés a aigua i bon clima, o de l'accés a energia barata. Totes aquests elements han estat motor de prosperitat, però ara en les condicions que ens trobem, el motor de la prosperitat és la capacitat de generar coneixement tècnic i científic contínuament. La capacitat d'innovar. Cada manera de sobreviure, demana, exigeix que els humans canviïn les seves motivacions per tal de poder respondre als reptes que presenta la nova manera de supervivència col·lectiva.

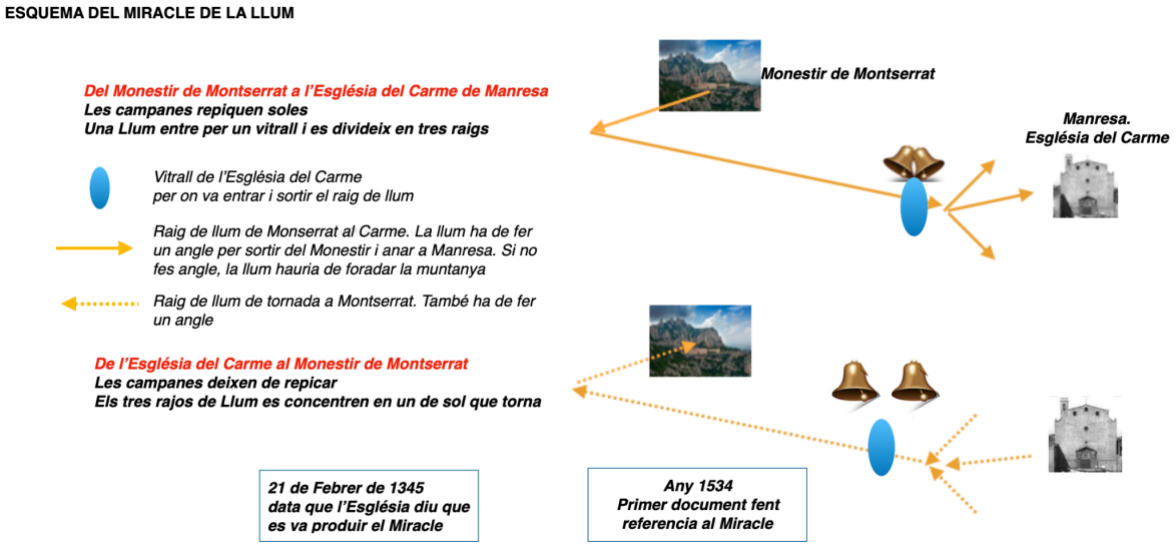
Per primera vegada en la història humana, la manera de sobreviure, aquesta necessitat d'innovació constant, ens demana entendre a nivell pràctic que el món que vivim és modelat, constituït semiòticament i per tant ens permet desenvolupar la capacitat per crear i innovar. Una de les nombroses conseqüències de la vida actual és que la pressió per innovar ens exigeix desenvolupar la capacitat de treballar en equip d'interdependència, i per tant que l'eix de l'èxit, no és l'individu ni la competència. Aquestes declaracions no són resultat d'un posicionament valoral, crític o moral, sinó una seqüela de la lògica d'innovar i del creixement accelerat de les ciències i les tecnologies. Per tant, defensar i comunicar una antropologia científica com a base de les formulacions teòriques, és a dir, que l'humà modela el món i està constituït pel llenguatge i no hipòtesis metafísiques com formulacions que el humà és cos i raó hauria de ser una noció cultural, compartida pels humans, per tal d'assegurar la prosperitat del col·lectiu. Una epistemologia de la ciència i de la vida no mítica, entenent que les formulacions científiques i les formulacions de la nostra vida quotidiana no descriuen la realitat sinó que son modelacions

[3] Tenim la física newtoniana, la física quàntica, amb diferents hipòtesis i models del món que funcionen.

adequades en una situació determinada és un dels ingredients fonamentals que permet les capacitats creatives humanes expressar-se i desenvolupar-se amb la màxima potència i diversitat. Les repercussions d'aquest canvis són monumentals.

Els treballs del Marià Corbí han permès explicar la font de la creativitat humana, explicar les configuracions axiològiques passades, i entendre els mites, les religions i les ideologies, sense quedar-nos atrapats en elements culturals ni mítics. Així, per primera vegada, en la història humana, podem plantejar-nos a nivell col·lectiu la orientació valoral humana com una construcció pensada, reflexionada, orientada i plenament conscient. Això és una revolució.

Clarifico que avui en dia hi ha encara religions, mites, ideologies i que hi ha molta població que continua sotmesa a aquests tipus de configuracions valorals. En el món occidental, però, aquestes configuracions ja no son totalitzadores, si no que les persones es sotmeten a elles de manera voluntària. Diguem que hi ha un "menú" on escollir. En el passat, això no era possible, cadascuna d'aquestes configuracions humanes era imposada, generalitzada a la comunitat a la que un pertanyia i assegurava la supervivència del col·lectiu. Per exemple, encara que avui en dia hi hagi molt bons cristians o musulmans la religió està desacreditada. El mateix passa amb les altres religions, amb les ideologies i els mites. Això no era així en el 1340 quan els pagesos de Manresa van construir una sèquia per poder regar els seus horts o morien de gana, el bisbe no volia que un menor cabal del riu fes que el seu molí, per moldre cereals, potser no treballés alguns mesos de l'any per falta de potència hídrica. Per tal d'aturar la construcció de la sèquia, el bisbe, va excomunicar a tota la ciutat, no es van celebrar cap ritus cristià durant més de cinc anys, va ser una crisi social de tal magnitud que només es va poder capgirar per l'arribada d'un nou bisbe que va rebre "un miracle" "de la llum" que venia del monestir de Montserrat i que va il·luminar el rosetó de l'església de Manresa -una fita totalment impossible per raons físiques-, però és clar, era un miracle, l'única manera de justificar que el parer del representat de Déu havia canviat.



Resum del Miracle de la Llum any 1345 - Manresa - visual creat per Pere Prat Puig

Aquella situació es recorda cada any després de quasi 700 anys, tal va ser l'efecte en generacions de manresans. En aquell temps tota la vida es vivia seguint els mandats de l'església, no es podia escapar de les configuracions axiològiques, i els manresans van haver de viure la complexitat d'un bisbe que estimava els diners més que la vida i feia uns discursos contra la gent i a favor del seu benefici material. Queda clar que només "un miracle" podia justificar revertir la situació.

Quan els individus poden decidir quina "religió seguir", com un menú en un restaurant, ja ens estan mostrant que les religions han deixat de ésser configuracions a nivell social. També ens permet definir la nostra situació col·lectiva actual com a desarticulada axiològicament i per tant molt diferent de com han estat configurades les poblacions humanes prèviament.

Vivim en societats sense una estructura axiològica col·lectiva. A nivell social s'ha comprès el fracàs dels projectes ideològics i les religions, fins i tot la ideologia dominant actual capitalista-neoliberal ha perdut reputació. Semblaria que no pot haver-hi lloc per noves formacions axiològiques, en "format" ideologia o religió, tenint en compte que la seva funcionalitat axiològica era adequada a una determinada manera de sobreviure[4].

Al mateix temps, la nostra comunitat global és un terreny fèrtil per multitud de noves propostes axiològiques, per donar resposta a les necessitats humanes d'orientació. Si entenem a l'humà con a constituït pel llenguatge, això vol dir que necessitem aquestes configuracions com l'aire que respirem. Per aquesta lògica axiològica avui dia tenim al nostra abast un "mercat" axiològic, amb multitud de creences, ideologies i religions. L'individu escull, dins d'aquestes possibilitats mítiques que li descriuen una veritat, una certesa. Sembla que aquest raonament que ofereixo porti a la creació d'un nou sistema totalitzador i únic. Però estic intentant explicitar que hem de crear un nou concepte axiològic, fugint de paraules i del que han estat els mites, religions i axiologies. Totes tenen en comú que pretenen descriure la realitat i orientar als humans de la manera correcta. Ho hem de transformar per tal de poder enfocar-nos en aquest repte que vol dir crear configuracions axiològiques col·lectives per la societat d'innovació continua que vivim, assumint plenament la incertesa i per tant consciència dels perills. El concepte que ha desenvolupat Corbí (2020) ha estat el de projectes axiològics col·lectius - on s'entén que les configuracions axiològiques són modelacions que s'han de construir.

Només, si a nivell col·lectiu es té plena consciència de les demandes axiològiques adequades per a la supervivència serem capaços de prosperar i les propostes seran no mítiques. Llavors, no serà l'individu qui es vegi forçat a escollir sinó que a nivell social, comunitari l'orientació vindrà guiada, compresa com a modelació. Com a societat encara som incapaços d'estructurar aquestes configuracions de valors orientades de manera sensitiva-racional connectades a com es sobreviu.

Les narratives dels creadors de la IA. La intel·ligència.

Novembre, 2022: data de naixement del Chat-GPT pel públic mundial. Amb un link a internet qualsevol persona podia interactuar amb un model de llenguatge desenvolupat per la empresa OpenAI. Aquests models de llenguatge entrenats amb tot el text d'internet i amb totes les transcripcions d'àudio i vídeo d'internet son capaços d'emular les converses humanes i per tant sembla que són intel·ligents.

El que anomenem Intel·ligència Artificial, IA, és un terme comercial que engloba molts tipus de tecnologies. La que està en boca de tots és la IA generativa, la que està a la base dels "chatbots", aquesta és la IA que està triomfant ara. És una tecnologia que no és funcional, no podem assegurar que les seves respostes, suggeriments, etc. siguin correctes, per tant per definició no és segura i per tant no podem assegurar que sigui efectiva ni tampoc eficient (Niederhoffer i altres, 2025).

OpenAI, al Novembre 2022 estava valorada en $29,000 milions i a Octubre 2025 està valorada en $ 500,000 milions (Kinder i Hammond, 2025), 16 vegades més. El seu valor ha augmentat estratosfèricament perquè han sabut vendre la història d'una manera efectiva que estan

[4] En el cas de les ideologies, són el resultat del procés de la industrialització.

construint una superintel·ligència. El chat-GPT és el primer pas, i segons el seu director general, Sam Altman, qui sigui capaç de fer-ho serà qui controli la humanitat i al mateix temps serà capaç de resoldre tots els nostres problemes[5]. Aquesta idea va ser descrita inicialment per IJGood (1966):
"Definim una màquina ultraintel·ligent com una màquina capaç de superar amb escreix totes les activitats intel·lectuals de qualsevol ésser humà, per molt enginyós que sigui. Atès que el disseny de màquines és una d'aquestes activitats intel·lectuals, una màquina ultraintel·ligent podria dissenyar màquines encara millors; això donaria lloc, de manera inqüestionable, a una "explosió d'intel·ligència", i la intel·ligència humana quedaria molt enrere. Així, la primera màquina ultraintel·ligent seria la darrera invenció que la humanitat necessitaria mai, sempre que fos prou dòcil per indicar-nos com mantenir-la sota control. És curiós que aquest punt s'hagi esmentat tan poques vegades fora de la ciència-ficció. De vegades val la pena prendre la ciència-ficció seriosament."

Aquest és un dels arguments claus per propulsar les inversions en intel·ligència artificial i els estudis dels riscos existencials d'aquesta tecnologia en detriment de l'enfoc en els problemes actuals que aquesta diversitat de tecnologies està causant.

Aquest pensament de l'intel·ligència com a producte és possible? la intel·ligència sempre és una avaluació respecte a un entorn dinàmic i sempre depèn d'uns perceptors, d'un cos. És a dir, la intel·ligència d'una formiga és diferent a la d'un humà i sempre està en relació a l'entorn. Com podem parlar d'una ultra-intel·ligència com si fos un producte, una entitat, quan és una resposta relacional i depenent d'unes dinàmiques específiques de supervivència? com podem "vendre" una ultra-intel·ligència que hem fabricat i "venem" com una intel·ligència comparable a la humana. No ho pot ser mai.

Open AI, com altres startups, Anthropic, Perplexity AI, MidJourney, segueixen les dinàmiques standards del "capital risc", son capaces de crear productes potencialment interessants però encara no hi ha un negoci, o no tenen ingressos o beneficis. Per tal d'aconseguir finançament, les startups presenten un futur, i els inversors "aposten" per una història, per una narrativa. S'aposta per la persona que és capaç de generar una narrativa creíble del futur, perquè el negoci no hi és encara. Es fa servir la metàfora que l'important és el jockey -l'impulsor, no el cavall - l'empresa. Si la persona és la correcta, sabrà trobar la manera de transformar o recrear la narrativa de tal manera que el negoci sigui un èxit. És per això que saber crear una història és el més important.

Així, es creen diverses històries per propulsar aquestes empreses. He fet un recull no exhaustiu, però al meu parer significatiu de les narratives més comunes.

La primera, "*el poder de la IA és tal que pot acabar amb la humanitat*[6]". Destacar el poder de la IA és una manera d'adquirir notorietat, de crear noves converses, discursos. Tenir-la sempre present a nivell públic. Els ciutadans, els polítics, les institucions i les empreses es veuen obligades a formular la seva posició. Això genera més probabilitats que els productes i serveis d'aquestes start-ups tinguin compradors.

Sembla contradictori que algú que vulgui vendre un producte/servei destaqui una part negativa del producte/servei, en aquest cas el "poder de la IA per acabar potencialment amb la humanitat". Els creadors de la IA presenten la possibilitat que pot ser utilitzada de manera criminal i s'ha de regular el seu ús. Però amb aquesta narrativa, no es qüestiona que simplement aquest producte/servei no hauria d'existir (Golumbia, 2022). O que aquest producte/servei no funciona, sinó que s'assumeixen que aquests dos aspectes (1) la seva legitimitat per ser

[5] https://blog.samaltman.com/

[6] https://www.fhi.ox.ac.uk/ - tancat al 2024 arrel d'una controvèrsia racista per part del seu director.

comercialitzada (2) la seva funcionalitat, i s'enfoca tota la reflexió en el ús negatiu de la tecnologia.

Ja se que això sembla molt poc lògic. M'agradaria comentar aquest exemple: https://ai-2027.com/. Els autors d'aquesta "recerca" defensen que és una recerca encara que es llegeix com un exercici de ciència "fictícia"[7]. Els autors argumenten que han abandonat carreres lucratives treballant per les start-ups de IA per dedicar-se a crear narratives com aquesta que avisen dels possibles futurs negatius que podem viure. Són "recercaires" que tenen consciència, que es preocupen pel benefici de la humanitat, són "els bons". També presenten altres credencials, per tal que ens creiem aquestes narratives, que ja han demostrat en el passat, amb anàlisis similars, que les seves prediccions són correctes o que han treballat en start-ups sobre el tema de seguretat de la IA.

En aquesta "creació" del futur que s'imaginen, s'ha de ressaltar que: **definir i preocupar-se pel futur crea poder en el present**, com passa això?: primer, es presenta el futur com a **resultat d'una inevitabilitat tècnica accelerada** que porta més d'hora o més tard a la creació d'una superintel·ligència, aquí no hi enlloc un reconeixement de tota la disciplina de ciència del coneixement ni de la interacció amb la societat i la política. Segon, degut a les hipòtesis axiològiques d'aquest grup de "recercaires", aquesta superintel·ligència és definida com dominant, colonial i exclusiva, *"qui controli aquesta superintel·ligència tindrà el poder"*.

Així, aquesta narrativa **ens impedeix discutir aquestes dues hipòtesi de base i ens centra, ens obliga, a pensar i reflexionar en les seves prediccions**. Des del meu punt de vista, es tracta d'un parany: quedem atrapats en una visió del futur en què la tecnologia és determinada pel capital risc, encara que ells ho presentin com si fos fruit de l'autonomia de la innovació científica. És evident que els autors passen per alt més de setanta anys d'estudis sobre el desenvolupament científic i tecnològic. Aquesta mirada sobre la IA manté com a "*lògica, racional i inevitable*" l'actual evolució de la IA, com si els interessos dominants no hi tinguessin cap influència. Un plantejament, com a mínim, sorprenent.

Hi ha molts investigadors que desenvolupen, el que ells anomenen, "*seguretat de la IA*", centrats en l'estudi i mitigació de riscos existencials - extinció de l'espècie humana-. La seva argumentació hereva de les teories filosòfiques utilitaristes de Peter Singer (1993) és seguit fervorosament per molts desenvolupadors de la IA als Estats Units. Un dels eixos d'aquesta comunitat és https://www.lesswrong.com/ i el seu enfoc en el *"longtermism"*, és a dir, en el efecte futur d'aquestes tecnologies (Torres, 2021).

Sembla estrany que, volent defensar el benefici de la humanitat em centri a criticar o atacar aquestes postures, que sembla també defensen el *"benefici de la humanitat"*: la seguretat de la IA, és un motiu lloable. Però el que realment estan fent és defensar una visió del futur on s'accepta la inevitabilitat d'una certa IA, i la "intel·ligència" com dominadora i explotadora i ens roben la possibilitat, fins i tot, d'imaginar un altre tipus de tecnologia.

Una altre de les narratives que circulen és: "*Nosaltres, els enginyers i programadors som molt intel·ligents, perquè som capaços de crear aquesta intel·ligència artificial, per això el que diem està carregat de raons i és molt intel·ligent i raonable*". Per aquesta raó, totes aquestes propostes són àmpliament amplificades, es pensen que els impulsors són molt intel·ligents, justificades per gent molt intel·ligent, al cap i a la fi són els titans del món! i per tant mereixen la seva amplificació en els mitjans tradicionals[8].

[7] No l'anomeno ciència ficció perquè no té com a objectiu entretenir creant futurs.

[8] https://www.nytimes.com/2025/04/11/podcasts/hardfork-tariffs-ai-2027-llama.html

Aquesta "intel·ligència" segueix els paràmetres valorals de persones molt específiques: els programadors, els enginyers, les persones que treballen en el capital risc a Silicon Valley.

El tema de la inevitabilitat de la IA també es fa servir com un argument de dominació geopolítica, *"Si no la creem nosaltres- els americans-, la crearan els Xinesos"*. Però aquesta *"inevitabilitat"* és una ficció, no és com la ciència i la tecnologia es desenvolupa.

Un altre aspecte de la inevitabilitat de la IA i de l'estudi de la *"seguretat de la IA"* és com es trasllada la mirada al futur i per tant, s'obliden i no es tenen en compte els problemes que actualment ja està generant la IA, per exemple, les xarxes socials generen múltiples problemes amb els seus sistemes recomanadors (molt poca intel·ligència), però que tenen un gran impacte.

Una altre narrativa dels titans actuals de la IA és la defensa desfermada del poder de la tecnologia per transformar el món i la suficiència moral i ètica d'aquesta posició. Qualsevol persona que s'oposi a aquesta visió es titllada d'ignorant i enemiga. No només es defensa aquest futur tecnològic sino que es defensa tota la racionalitat. El mercat es la manera més racional de prendre decisions. La política és un destorb. La burocràcia s'ha de destruir[9]. En definitiva, una defensa del capital risc per tal que no tingui cap mena d'entrebanc i pugui continuar funcionant. Tot en *"benefici de la humanitat"*. Aquí es defensa el manteniment i aprofundiment de l'economia capitalista neoliberal de mercat. L'objectiu d'aconseguir desenvolupar aquesta intel·ligència artificial és tan important que els problemes que poden passar pel fet d'intentar aconseguir aquest objectiu no són rellevants: problemes climàtics, socials, geopolítics, guerres, etc.

Segons aquesta narrativa tampoc ens hem de preocupar dels problemes que estem creant amb el desenvolupament actual de la IA, per exemple: "*l'explotació humana en la creació de dades* (O'Neil, 2017; Gillespie, 2018; Crawford, 2021), *la automatització de la discriminació* (Eubanks, 2019; Benjamin, 2019; Buolamwini, 2023, Broussard, 2023), *l'explotació de la natura* (Crawford, *2021) una vegada aconseguida la fita d'una intel·ligència artificial general, llavors resoldrem tots els problemes, com per exemple la crisis climàtica*". Sense reconèixer que la gran majoria dels "problemes" humans no són tecnològics, sinó que són socials, culturals. Si, per exemple, tinguéssim una IA que fos un oracle, tipus com se l'imagina Elon Musk[10], no resoldríem els problemes socials i culturals. Per exemple, ara tenim el coneixement per resoldre la crisis climàtica però no hem desenvolupat la voluntat política i social.

Els "gurus" de la IA han creat histories impactants explicitant les distopies de control de la IA sobre els humans, i que per tant s'han de regular i evitar l'extinció de la humanitat, però quan hi ha regulacions de la IA, com la recent AI Act (2024) de la Unió Europea, totes les empreses americanes que controlen la frontera del desenvolupament de la IA hi estan en contra, de fet, l'administració del Trump està pressionant per aturar la regulació[11]. Sembla que Europa està endarrerint la implementació de la llei, AI Act.

La intel·ligència com aspecte definidor del que és ser humà comparat amb els animals també s'ha fet servir per "vendre" aquest nou producte/servei. La IA és "intel·ligent", però és una intel·ligència peculiar, la que ara domina, la IA generativa, es una mirada estadística al món, no te res d'intel·ligència humana, capaç de comprendre qualitats, fer abstraccions, valorar les situacions, etc. La IA generativa emula la intel·ligència humana sense funcionar com els humans ho fan.

[9] https://a16z.com/the-techno-optimist-manifesto/

[10] https://abcnews.go.com/Business/elon-musk-launches-ai-company-compete-chatgpt/story?id=101210078

[11] Haeck, P. (2025, June). EU's waffle on artificial intelligence law creates huge headache. *Politico*.

S'ha destacat aquest aspecte, la seva "intel·ligència", perquè ajudava de manera important a fer créixer l'interès per productes i serveis que fossin intel·ligents. Aquesta intel·ligència es destaca de moltes maneres, normalment justificant que el model de IA és capaç de passar molts tests, diferents tipus de tests. De fet, hi ha una pàgina web (Language Models Arena) que mitjançant la participació dels usuaris (https://lmarena.ai/) fa un ranking dels millors models de llenguatge.

Els rankings que es creen, a partir de "benchmarks", és a dir, especificacions tècniques que avaluen aspectes essencials dels models, però que per ser efectius en l'avaluació dels models han de ser concrets i per tant no son efectius per donar una apreciació del nivell d'intel·ligència del sistema de IA. Per exemple, el fet que un model de IA sigui capaç de respondre l'examen per ser advocat, no vol dir que pugui ser un bon advocat, hi ha molt més en la feina d'advocat que tenir coneixement, s'ha de saber d'estratègia, coneixement social, etc.

Per tant, és una "intel·ligència" diguem, molt específica, la capacitat per contestar tests. A més, per encara fer-ho més poc comparable a la intel·ligència humana, els creadors dels models no poden assegurar que les dades dels exàmens que fan els models de llenguatge no s'han fet servir per entrenar el model[12]. També aquests models "raonen" de maneres molt específiques, millor dit, no son capaços de raonar[13].

Aquesta idea que la IA ens proporcionarà intel·ligència i per tant tots els nostres problemes es podran solucionar te unes hipòtesis dolentes a nivell epistemològic i també a nivell social. Hi ha molt pocs argumentant contra aquestes idees (Marcus, 2024; Bender i Hanna, 2025, McQuillan, 2022).

Científics, com Geoffrey Hinton, que es dediquen a la computació fan declaracions més enllà del seu camp d'expertesa (Lohr, 2025) i pel fet de ser científics de la computació, de haver guanyat nobels o de ser directors generals de aquestes start-ups tecnològiques el que diuen és considerat com de molt valor. Per exemple, el Elon Musk[14] defensa que la IA ens donarà *"la veritat"*, com si fos un oracle. Aquest és un error epistemològic. L'humà modela la realitat i per tant, per definició no hi ha "la veritat", a la que un ésser amb suficient intel·ligència pot accedir, sino modelacions que són més o menys efectives i que van canviant segons les nostres necessitats. Segurament, aquest tipus de mirada estadística que ens dona la intel·ligència artificial ens ajudarà a crear modelacions més efectives, però la tecnologia que tenim ara sempre està basada en dades generades per humans, i per tant amb uns perceptors limitats i unes màquines limitades que amplifiquen les nostres capacitats, sempre per definició tindrem límits. La modelació és infinita i els límits també. La nostra estructura biològica que requereix una constitució per ser viable a través del llenguatge ens ha ofert aquesta mirada única en el misteri que vivim. Pensar que la IA ens revelarà *"la veritat"* és una recepta pel fonamentalisme i la violència legitimada per una intel·ligència "superior" inexistent.

Hauríem de ser capaços de sortir d'aquest tipus de narratives, la IA tindrà impacte depenent de qui la fa servir, com la fa servir i com està imbricada en sistemes de influència o control (Mühlhoff, 2025). Per exemple, des de fa més de dues dècades les xarxes socials fan servir sistemes recomanadors, que és un tipus d'algorisme, diríem de IA, no gaire intel·ligent, per standards humans, però si amb molt impacte. Per exemple, empreses com Uber el fan servir per crear els seus serveis de transport i crear precarietat laboral, empreses com Meta venen els seus serveis i la informació que han recopilat dels seus usuaris per desenvolupar campanyes de

[12] Si un model ha estat entrenat per exemple amb les solucions a preguntes d'exàmens, llavors una vegada entrenats si se'ls hi fan aquelles preguntes seran capaços de contestar-les sense problemes.

[13] Horton, M. (2025). *The Illusion of Thinking : Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity*. 1–30.

[14] https://abcnews.go.com/Business/elon-musk-launches-ai-company-compete-chatgpt/story?id=101210078

manipulació política, empreses com Palantir venen serveis de vigilància basats en dades públiques, etc.

La conversa actual sobre la intel·ligència de la IA està centrada en enfocar la nostra atenció en un problema tècnic "com fer que la IA sigui intel·ligent", però l'impacte de la IA no dependrà de si resolem aquest "problema tècnic" sinó com fem servir aquesta IA a la nostra societat, qui la controla i com està imbricada en sistemes de control i influència. Per tant, s'ha d'obrir l'espai per tal que disciplines que tenen a veure amb els aspectes humans (antropologia, axiologia, sociologia, psicologia, humanitats, filosofia i economia) ajudin a definir quin tipus de IA volem. I aquestes reflexions no poden sortir dels creadors d'aquests sistemes "intel·ligents" perquè amb les seves capacitats i orientacions específiques i limitades perpetuen una mirada tècnica. Per tant, son incapaços de aportar solucions i repliquen, aprofundeixen i automatitzen els problemes actuals.

Les pràctiques dels creadors de la IA, un exemple: Google

M'agradaria presentar l'exemple de la Timnit Gebru per il·lustrar com actúen més enllà de les narratives que impulsen el sector tecnològic dels Estats Units i què vol dir l'ètica de la IA pels creadors de l'IA, en aquest cas a Google.

La Dra.Gebru és una científica especialitzada en intel·ligència artificial que des del 2018 treballava a Google com a co-líder de l'equip d'IA ètica juntament amb Margaret Mitchell. L'any 2020 va escriure, conjuntament amb ella i altres autors, un article titulat "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" per a la conferència ACM Conference on Fairness, Accountability, and Transparency, en què alertava sobre els riscos mediambientals i la perpetuació de biaixos en els models de llenguatge a gran escala, especialment en sistemes com BERT de Google (entrenat amb 3,3 mil milions de paraules) o GPT-3 (amb mig bilió de paraules), perquè aquests models fan servir grans quantitats de dades que no han estat "curades" i per tant tenen tots els biaixos i les problemàtiques presents en la nostra societat **es perpetuen de manera automatitzada.**

Aquest article, que ja havia passat una primera revisió interna, va ser bloquejat per la direcció de Google, al·legant que era massa crític i no contemplava prou els mecanismes de mitigació de riscos. El 2 de desembre de 2020, després que Timnit Gebru demanés explicacions sobre el procediment de censura interna, va rebre un correu electrònic que "acceptaven la seva renúncia", malgrat que ella mai no havia presentat cap dimissió. Aquest fet va desencadenar una forta polèmica: prop de tres mil empleats de Google i més de quatre mil membres de la comunitat acadèmica i civil van firmar un manifest a favor seu. Gebru, que ja era reconeguda per la seva recerca sobre biaixos algorítmics, com ara l'estudi *"Gender Shades"* conjuntament amb Buolamwini (2018), on va evidenciar els elevats errors de reconeixement facial en dones negres, també havia denunciat que Google no tenia una política efectiva de promoció de la diversitat, i va escriure en un correu intern que *"no hi ha cap incentiu real per contractar més dones o persones subrepresentades"*.

Davant la protesta, el director general de Google, Sundar Pichai, va enviar una disculpa interna dient: *"He sentit la reacció davant la marxa de la Dra. Gebru alt i clar"* i prometent un procés de revisió, però molts treballadors ho van considerar insuficient. Poc després, Google va acomiadar també Margaret Mitchell. El 2 de desembre de 2021, just un any després dels fets, Timnit Gebru va fundar el Distributed AI Research Institute (DAIR)[15], amb l'objectiu de fomentar una recerca en intel·ligència artificial ètica, descentralitzada i respectuosa amb les comunitats més afectades per la tecnologia.

[15] https://www.dair-institute.org/

Aquesta situació, ens planteja multitud de preguntes. Veritablement és possible un rol d'ètic de la IA en una empresa tecnològica? perquè ens estem enfocant en una ètica quan aquests productes "intel·ligents" potser no haurien ni d'existir? és un rol d'aquestes característiques -ètiques- una feina que pot tenir un impacte rellevant en com es desenvolupen aquestes tecnologies? perquè hi han aquest tipus de rols en les empreses tecnològiques? poden equips ètics funcionar de manera efectiva a les organitzacions? Tenint en compte el que sabem sobre els problemes que presenten el models de llenguatge que son la base de productes com el Chat GPT (OpenAI), Claude (Anthropic) o Gemini (Google), per exemple, problemes amb repetir i amplificar biaixos, creació eficaç de desinformació, còpia de continguts amb copyright (Hammond i Acton, 2025) falta d'explicabilitat dels resultats, facilitat per fabricar informació errònia (al·lucinacions), la despesa energètica exorbitant etc. quina funció tenen aquests departaments d'ètica per la IA generativa?

A l'octubre de 2024, un nen de 14 anys (Montgomery, 2024), perdudament enamorat d'un personatge de ficció amb qui mantenia una relació mitjançant un servei de chatbots[16], es va suicidar. A la base d'aquest servei hi ha els models de llenguatge que tots nosaltres fem servir com ChatGPT o Gemini, però especialitzats, en aquest cas mostrant unes característiques determinades, basades en el personatge Daenerys Targaryen de Joc de Trons. El problema, però, no és només tècnic, sinó que te a veure com aquests productes i serveis interaccionen amb els humans i les comunitats que els fan servir. La mare del Sewell ha demandat a la empresa que continua oferint una plataforma de compartició on hi ha més de 1000 personatges[17] diferents per tal que l'usuari triï amb quin interaccionar. N'hi ha de tots tipus, fins i tot abusadors. Hi ha molts serveis similars, n'hi ha que presenten com *"companys"*[18] amicals o sexuals per establir relacions enfocades a les necessitats individuals del subscriptor del servei. Però fins i tot el ChatGPT pot facilitar la mort d'un adolescent de 16 anys, tal com va passar a l'abril del 2025 (Hill, 2025).

Conclusió

Des de l'aparició de les xarxes socials hem vist com algoritmes cada vegada més sofisticats son capaços, segons els seus creadors, de *"connectar-nos"*, *"establir relacions"* i en general *"viure millor"*. Només considerant el sistema d'IA generativa, ChatGPT, més de 700 milions de persones envien al sistema més de 18.000 milions de missatges cada setmana (Clark i Nevitt, 2025). Ara aquests algorismes són molt *"intel·ligents"* i poden *"resoldre"* encara més problemes. Però i si totes aquestes avantatges fossin realment un parany? l'etapa inicial d'un món on és impossible veritablement entendre als altres i entendre el nostre món? això és el que argumenta el Centre per Tecnologia Humana[19]. Altres investigadors com Zuboff (2022) descriuen com aquestes empreses amb els seus productes i les seves promeses *"per connectar-nos"* o per *"buscar"* a la web, en realitat estan construint constantment perfils de qui som per tal de fer-nos més predictibles en les nostres actuacions. Vendre aquest coneixement a anunciants i així ser capaços de proveir-nos amb els millors productes/serveis pels quals paguem.

Així, per exemple en el cas de les xarxes socials, la promoció d'informació esbiaixada i polaritzant fa que les persones estiguin més interessades i per tant estan més temps fent servir aquests serveis, destruint, pel benefici financer d'aquestes empreses, la comprensió del món que tenim, i per tant l'espai comú per la gestió política de la nostra comunitat. Tot per vendre aquestes dades a anunciants i fer-nos més predictibles. Cada vegada que fem servir Google, els

[16] https://character.ai/

[17] https://character.ai/sitemap/characters_a

[18] https://replika.com/

[19] https://www.humanetech.com/

resultats inicials que ens donin dependrà del perfil que tenen els algoritmes de qui som i què pot ser el nostre interès. Per tant, ens estan presentant un món totalment centrat en com l'algoritme ens ha valorat, i d'això també dependrà el preu amb el que paguem els productes (Crawford, 2025).

Pot la IA amb la seva *"intel·ligència"* aconseguir-nos aquest món que els seus creadors prometen? Com pot una mirada ètica de la IA pot incidir en aquestes dinàmiques competitives?. No estem creant una ficció, pensant que crearem una ètica que resoldrà aquest problemes? L'objectiu comercial és que els usuaris estiguin el màxim del seu temps connectats en aquest serveis per tal que puguin deixar el màxim de dades sobre els seus interessos i que en posterior interaccions, aquestes empreses puguin vendre aquesta informació dels usuaris a altres com per exemple anunciants, però també persones interessades en afectar les opinions polítiques, com va ser el cas de Cambridge Analytica (Wylie, 2019), una empresa que s'ha demostrat, mitjançant una relació estreta amb Facebook, ara anomenada Meta, va ser capaç d'evitar que molts votants demòcrates anessin a votar i facilitant d'aquesta manera l'elecció de Donald Trump al 2016.

Enfocar-nos en desenvolupar una ètica de la IA, en les condicions actuals de la nostra comprensió del què és una ètica de la IA, ens porta a un carreró sense sortida (Munn, 2022; Phan i altres, 2022). Per què continuem creient que l'ètica de la IA pot solucionar les dinàmiques explotadores del nostre sistema econòmic? o que les empreses de IA amb la seva idea de beneficiar als humans veritablement es preocupen pels efectes dels seus productes i serveis en els humans que els fan servir?. Per què imaginem o explicitem una ètica optativa que no va en consonància amb l'ètica real, la de veritat, que es destil·la del projecte valoral actual? Per què pensem que actituds individuals voluntàries-basades en una ètica teòrica fictícia- poden contrarrestar el funcionament del sistema que té una altra ètica?.

Què pot fer un departament d'ètica de la IA, en una empresa tecnològica, en un entorn enfocat a aconseguir que l'usuari passi el màxim de temps possible amb el servei, sinó fer veure que és té cura de l'usuari? L'empresa defensa que protegeix els interessos de l'usuari, però en realitat, això no és així, l'empresa té un objectiu comercial i treu un rendiment directe en relació al temps que l'usuari fa servir el servei per tal de vendre les seves dades.

Les empreses que comercialitzen la IA generativa o les xarxes socials, amb els seus algorismes recomanadors, defensen que els usuaris no estan sent obligats a fer-los servir sinó que escolleixen lliurement. El fet que sigui una elecció lliure vol dir que per definició és una elecció que beneficia a qui la fa? Els sistemes estan creats per fer la vida fàcil i convenient per tal que el individus acceptin invertir el seu temps en aquests sistemes de IA, el benefici se'l emporten les empreses i els costos i problemes que generen aquests sistemes la societat (Mühlhoff, 2025).

Quan un sistema de IA, com per exemple, el chatGPT es llança al mercat, com a mínim, hauria de respectar les lleis actuals, cap va respectar cap dret a la propietat intel·lectual (Metz i altres, 2024), ni han hagut de demostrar que són segures (McCabe, 2024), ni són transparents en els recursos que fan servir, ni com gestionen aquestes dades privades (Magid, 2025). S'han creat i s'estan creant contínuament tota una sèrie de riscos que la societat ha d'assumir. Fins ara aquestes empreses no han assumit cap responsabilitat (O'Neil, 2017; Olson, 2024; Schellmann, 2024; Wynn-Williams, 2025).

Penso que el problema que tenim és greu, continuem replicant respostes que no han aportat solucions, em refereixo a més de 15 anys amb els algorismes recomanadors de les xarxes socials: enfocant-nos en injustícies puntuals i aproximacions ètiques que creen la il·lusió que estem fent alguna cosa. Hauríem de ser capaços de poder incidir en el desenvolupament de la IA a la societat, no el desenvolupament tècnic, sinó el de la imbricació dels sistemes de IA a la societat,

tots i cadascun de nosaltres en la mesura que ens afecten. Com deia Dewey (1927), en les verdaderes democràcies, els ciutadans són l'única autoritat legítima, perquè els seus problemes creen el marc en què pot funcionar tota expertesa. Ara mateix no sembla que ens puguem imaginar futurs compartits per fer servir les tecnologies cada vegada més poderoses que estem creant pel benefici de la humanitat i la vida en general. Què estem esperant?

Referències:

Bender, E. M., & Hanna, A. (2025). *The AI Con*. Harper Collins.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *FAccT 2021 - Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.

Benjamin, R. (2019). *Race after technology*. Polity.

Bloor, D. (1991). *Knowledge and social imagery*. University of Chicago Press.

Broussard, M. (2023). More than a Glitch. In *More than a Glitch*. MIT Press.

Buolamwini, J. (2023). *Unmasking AI : my mission to protect what is human in a world of machines*.

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*, *81*, 77–91.

Callon, M. (2010). Afterword. In A. Feenberg (Ed.), *Between reason and experience: Essays in technology and modernity*. The MIT Press.

Clark, D., & Nevitt, C. (2025, October 7). How AI became our personal assistant. *Financial Times*.

Corbí, M. (1983). *Análisis epistemológico de las configuraciones axiológicas humanas. La necesaria relatividad cultural de los sistemas de valores humanos: mitologías, ideologías, ontologías y formaciones religiosas.* Ediciones Universidad de Salamanca

Corbí, M. (2020). *Proyectos colectivos para sociedades dinámicas*. Herder.

Crawford, D. (2025). *Surveillance pricing: How your data determines what you pay*.

Crawford, K. (2021). *Atlas of AI*. Yale University Press.

De Saussure, F. (1959). *Course in general linguistics*. Columbia University Press.

Dewey, J. (1927). *The public and its problems*. H. Holt and Company.

Eubanks, V. (2019). *Automating inequality : how high-tech tools profile, police, and punish the poor*. Picador.

European Parliament, & European Council. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024. *Official Journal of the European Union*, *1689*(3), 1–144. http://data.europa.eu/eli/reg/2024/1689/oj

Gillespie, T. (2018). *Custodians of the internet*. Yale University Press.

Golumbia, D. (2022, December). ChatGPT Should Not Exist. *Medium*.

Good, I. J. (1966). *Speculations Concerning the First Ultraintelligent Machine*. *May*, 31–88.

Haeck, P. (2025, June). EU's waffle on artificial intelligence law creates huge headache. *Politico*.

Hammond, G., & Acton, M. (2025, September 5). AI start-up Anthropic settles landmark copyright suit for $1.5bn. *Financial Times*.

Haraway, D. (1988). Situated Knowledges: the Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, *14*(3), 575–599.

Harding, S. (2017). *Whose Science? Whose Knowledge?*

Hill, K. (2025). A Teen Was Suicidal. ChatGPT Was the Friend He Confided In. *New York Times*.

Horton, M. (2025). *The Illusion of Thinking : Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity*. 1–30.

Kinder, T., & Hammond, G. (2025, October 2). OpenAI overtakes SpaceX after hitting $ 500bn valuation. *Financial Times*.

Kuhn, T. S. (1970). *The structure of scientific revolutions*. The University of Chicago Press.

Latour, B. (2005). *Reassembling the social : an introduction to actor-network-theory* (Repr.). Oxford Univ. Press.

Law, J. (2017). STS as Method. In U. Felt (Ed.), *The handbook of science and technology studies*.

Lohr, S. (2025). Your A.I. Radiologist Will Not Be With You Soon. *New York Times*.

Magid, Y. (2025). *AppleStorm: Unmasking the Privacy Risks of Apple Intelligence*. Lumia.

Marcus, G. (2024). *Taming Silicon Valley*. The MIT Press.

McCabe, D. (2025). Regulators Are Digging Into A.I. Chatbots and Child Safety. *New York Times*.

McQuillan, D. (2022). Resisting AI. In *Resisting AI*. Bristol University Press.

Metz, C., Kang, C., Fenkel, S., Thompson, S., & Grant, N. (2024). How Tech Giants Cut Corners to Harvest Data for A.I. *New York Times*.

Montgomery, B. (2024). Mother says AI chatbot led her son to kill himself in lawsuit against its maker | Artificial intelligence (AI). *The Guardian*, 11–13.

Mühlhoff, R. (2025). *The Ethics of AI. Power, Critique , Responsibility*. Bristol University Press.

Munn, L. (2022). The uselessness of AI ethics. *AI and Ethics*.

Niederhoffer, K., Kellerman, G. R., Lee, A., Liebscher, A., Rapuano, K., & Hancock, J. T. (2025). AI-Generated " Workslop " Is Destroying Productivity. *Harvard Business Review*.

O'Neil, C. (2017). *Weapons of math destruction : how big data increases inequality and threatens democracy*. Penguin Books.

Olson, P. (2024). *Supremacy*. St. Martin's Press.

Phan, T., Goldenfein, J., Mann, M., & Kuch, D. (2022). Economies of Virtue: The Circulation of 'Ethics' in Big Tech. *Science as Culture*, *31*(1), 121–135.

Shapin, S. (1984). Pump and circumstance: Robert Boyle's literary technology. *Social Studies of Science*, *14*, 481–520.

Singer, P. (1993). *Practical Ethics*. Cambridge University Press.

Torres, P. (2021). The Dangerous Ideas of "Longtermism" and "Existential Risk." *Current Affairs*, 1–22.

Wylie, C. (2019). *Mindf*ck : Cambridge Analytica and the plot to break America*. Random House.

Wynn-Williams, S. (2025). *Careless people*. Flatiron Books.

Zahn, M. (2023). *Elon Musk launches his own AI company to compete with ChatGPT*. https://abcnews.go.com/Business/elon-musk-launches-ai-company-compete-chatgpt/story?id=101210078