

INTEL·LIGÈNCIA ARTIFICIAL I PAU

**Missatge del Sant Pare Francesc
per a la 57a Jornada Mundial de la Pau**

1 de gener de 2024

Al començament del nou any, temps de gràcia que el Senyor ens dona a cadascú de nosaltres, voldria adreçar-me al Poble de Déu, a les nacions, als caps d'estat i de govern, als representants de les diverses religions i de la societat civil, i a tots els homes i dones del nostre temps per expressar-los els meus desitjos de pau.

1. El progrés de la ciència i de la tecnologia com a camí vers la pau

La Sagrada Escripura atesta que Déu ha donat als homes el seu Esperit perquè tinguin «talent, intel·ligència i coneixements, i sigui un expert en tota mena de treballs» (Ex 35,31). La intel·ligència és expressió de la dignitat que ens ha donat el Creador en fer-nos a imatge i semblança seva (cf. Gn 1,26) i ens ha fet capaços de respondre al seu amor a través de la llibertat i del coneixement. La ciència i la tecnologia manifesten de manera particular aquesta qualitat fonamentalment relacional de la intel·ligència humana: ambdues són producte extraordinari del seu potencial creatiu.

A la Constitució pastoral *Gaudium et spes*, el Concili Vaticà II ha insistit en aquesta veritat, declarant que «l'home s'ha esforçat sempre amb el treball i l'enginy a desenvolupar més i més la seva vida»¹. Quan els éssers humans, «amb l'ajut de la tècnica», s'esforcen perquè la terra «esdevingui un estatge digne de tota la família humana»², actuen segons el disseny de Déu i cooperen amb la seva voluntat de dur a terme la creació i de difondre la pau entre els pobles. També el progrés de la ciència i de la tècnica, en la mesura que contribueix a un millor ordre de la societat humana i a augmentar la llibertat i la comunió fraterna, porta al perfeccionament de l'home i a la transformació del món.

Ens n'alegrem justament i estem agraïts per les extraordinàries conquestes de la ciència i de la tecnologia, gràcies a les quals s'ha pogut posar remei a innumbrables mals que afligien la vida humana i causaven grans patiments. Al mateix temps, els progressos tecnicocientífics, fent possible l'exercici d'un control sobre la realitat, mai vist fins ara, estan posant en mans de l'home una àmplia gamma de possibilitats, algunes de les quals poden representar un risc per a la supervivència humana i un perill per a la casa comuna³.

Els notables progressos de les noves tecnologies de la informació, especialment a l'esfera digital, presenten, per tant, apassionants oportunitats i riscos greus, amb serioses implicacions per a la recerca de la justícia i de l'harmonia entre els pobles. Per tant, cal plantejar-se algunes preguntes urgents. Quines seran les conseqüències, a mig i a llarg termini, de les noves tecnologies digitals?

¹ N. 33.

² *Ibid.*, n. 57.

³ Cf. Carta enc. *Laudato si'* (24 de maig de 2015), 104.

I quin impacte tindran sobre la vida dels individus i de la societat, sobre l'estabilitat internacional i sobre la pau?

2. *El futur de la intel·ligència artificial entre promeses i riscos*

Els progressos de la informàtica i el desenvolupament de les tecnologies digitals els últims decennis ja han començat a produir transformacions profundes en la societat global i en les seves dinàmiques. Els nous instruments digitals estan canviant el rostre de les comunicacions, de l'administració pública, de la instrucció, del consum, de les interaccions personals i d'altres innombrables aspectes de la vida quotidiana.

A més, les tecnologies que utilitzen un gran nombre d'algoritmes poden extreure, dels rastres digitals deixats a Internet, dades que permeten controlar els hàbits mentals i relacionals de les persones amb finalitats comercials o polítiques, sovint sense que ho sàpiguen, limitant-los l'exercici conscient de la llibertat d'elecció. De fet, en un espai com el web, caracteritzat per una sobrecàrrega d'informacions, es pot estructurar el flux de dades segons criteris de selecció no sempre percebuts per l'usuari.

Hem de recordar que la recerca científica i les innovacions tecnològiques no estan desencarnades de la realitat ni són «neutrals»⁴, sinó que estan subjectes a les influències culturals. En tant que activitats plenament humanes, les direccions que prenen reflecteixen decisions condicionades pels valors personals, socials i culturals de cada època. El mateix s'ha de dir dels resultats que aconseguixen. Aquests, precisament en tant que fruit d'acostaments específicament humans al món circumstant, tenen sempre una dimensió ètica, estretament lligada a les decisions de qui projecta l'experimentació i enfoca la producció vers objectius particulars.

Això val també per a les formes d'intel·ligència artificial, per a la qual, fins avui, no hi ha una definició unívoca en el món de la ciència i de la tecnologia. El mateix terme, que ja ha entrat en el llenguatge comú, abraça una varietat de ciències, teories i tècniques dirigides a fer que les màquines reproduïxin o imitin, en el seu funcionament, les capacitats cognitives dels éssers humans. Parlar en plural de “formes d'intel·ligència” pot ajudar a subratllar sobretot la diferència que no es pot omplir que existeix entre aquests sistemes, per més sorprenents i potents que siguin, i la persona humana. Aquests són, al cap i a la fi, “fragmentaris”, en el sentit que només poden imitar o reproduir algunes funcions de la intel·ligència humana. L'ús del plural posa en evidència a més que aquests dispositius, molt diferents entre ells, s'han de considerar sempre com a “sistemes socio-tècnics”. En efecte, el seu impacte, independentment de la tecnologia de base, no només depèn del projecte, sinó també dels objectius i dels interessos de qui els posseeix i de qui els desenvolupa, així com de les situacions en les que s'usen.

La intel·ligència artificial, per tant, ha de ser entesa com una galàxia de realitats diferents i no podem presumir *a priori* que el seu desenvolupament aportí una contribució benèfica al futur de la humanitat i a la pau entre els pobles. Aquest resultat positiu només serà possible si som capaços d'actuar de forma responsable i de respectar els valors humans fonamentals com «la inclusió, la transparència, la seguretat, l'equitat, la privacitat i la confiança»⁵.

No n'hi ha prou ni tan sols amb pressuposar, de part de qui projecta algoritmes i tecnologies digitals, un compromís d'actuar de forma ètica i responsable. És precís reforçar o, si cal, instituir

⁴ Cf. *ibid.*, 114.

⁵ *Discurs als participants a la trobada “Minerva Dialogues”* (27 de març de 2023).

organismes encarregats d'examinar les qüestions ètiques emergents i de tutelar els drets dels que utilitzen formes d'intel·ligència artificial o en reben la influència⁶.

La immensa expansió de la tecnologia, per tant, ha de ser acompanyada, per al seu desenvolupament, per una adequada formació en la responsabilitat. La llibertat i la convivència pacífica estan amenaçades quan els éssers humans cedeixen a la temptació de l'egoisme, de l'interès personal, de l'afany de lucre i de la set de poder. Tenim per això el deure d'ampliar la mirada i d'orientar la recerca tècnico-científica cap a la consecució de la pau i del bé comú, al servei del desenvolupament integral de l'home i de la comunitat⁷.

La dignitat intrínseca de cada persona i la fraternitat que ens vincula com a membres d'una única família humana, han d'estar a la base del desenvolupament de les noves tecnologies i servir com a criteris indiscutibles per a valorar-les abans del seu ús, de manera que el progrés digital pugui realitzar-se en el respecte de la justícia i contribuir a la causa de la pau. Els desenvolupaments tecnològics que no porten a una millora de la qualitat de vida de tota la humanitat, sinó que, al contrari, agreugen les desigualtats i els conflictes, no podran ser considerats un veritable progrés⁸.

La intel·ligència artificial serà cada vegada més important. Els reptes que planteja no són només tècnics, sinó també antropològics, educatius, socials i polítics. Promet, per exemple, un estalvi d'esforços, una producció més eficient, transports més àgils i mercats més dinàmics, a més d'una revolució en els processos de recopilació, organització i verificació de les dades. És necessari ser conscients de les ràpides transformacions que estan passant i gestionar-les de manera que es puguin salvaguardar els drets humans fonamentals, respectant les institucions i les lleis que promouen el desenvolupament humà integral. La intel·ligència artificial hauria d'estar al servei del millor potencial humà i de les nostres aspiracions més altes, no en competència amb ells.

3. La tecnologia del futur: màquines que aprenen soles

En les seves múltiples formes la intel·ligència artificial, basada en tècniques d'aprenentatge automàtic (*machine learning*), encara que es trobi en una fase pionera, ja està introduint canvis notables en el teixit de les societats, exercint una profunda influència en les cultures, en els comportaments socials i en la construcció de la pau.

Desenvolupaments com el *machine learning* o com l'aprenentatge profund (*deep learning*) plantegen qüestions que transcendeixen els àmbits de la tecnologia i de l'enginyeria i tenen a veure amb una comprensió estretament connectada amb el significat de la vida humana, els processos bàsics del coneixement i la capacitat de la ment d'assolir la veritat.

L'habilitat d'alguns dispositius per a produir textos sintàcticament i semàntica coherents, per exemple, no és garantia de fiabilitat. Es diu que poden "al·lucinar", és a dir, generar afirmacions que a primera vista semblen plausibles, però que en realitat són infundades o delaten prejudicis. Això planteja un seriós problema quan la intel·ligència artificial s'utilitza en campanyes de desinformació que difonen notícies falses i porten a una creixent desconfiança vers els mitjans de comunicació. La confidencialitat, la possessió de les dades i la propietat intel·lectual són altres

⁶ Cf. *ibid.*

⁷ Cf. *Missatge al President Executiu del "World Economic Forum" a Davos-Klosters* (12 de gener de 2018).

⁸ Cf. Carta enc. *Laudato si'*, 194; *Discurs als participants al Seminari sobre "El bé comú a l'era digital"* (27 de setembre de 2019).

àmbits en els que les tecnologies en qüestió plantegen greus riscos, als que s'afegeixen ulteriors conseqüències negatives unides al seu ús impropï, com la discriminació, la interferència en els processos electorals, la implantació d'una societat que vigila i controla les persones, l'exclusió digital i la intensificació d'un individualisme cada vegada més desvinculat de la col·lectivitat. Tots aquests factors corren el risc d'alimentar els conflictes i d'obstaculitzar la pau.

4. *El sentit del límit en el paradigma tecnocràtic*

El nostre món és massa ampli, variat i complex per a poder ser completament conegut i classificat. La ment humana mai no podrà esgotar-ne la riquesa, ni tan sols amb l'ajuda dels algorismes més avançats. Aquests, de fet, no ofereixen previsions garantides del futur, sinó només aproximacions estadístiques. No tot pot ser pronosticat, no tot pot ser calculat; al final «la realitat és superior a la idea»⁹ i, per més prodigiosa que pugui ser la nostra capacitat de càlcul, hi haurà sempre un residu inaccessible que escapa a qualsevol intent de quantificació.

A més, la gran quantitat de dades analitzades per les intel·ligències artificials no és de per sí garantia d'imparcialitat. Quan els algorismes extrapolen informacions, sempre corren el risc de distorsionar-les, reproduint les injustícies i els prejudicis dels ambients en els que s'originen. Com més veloços i complexos es tornen, més difícil és comprendre per què han generat un determinat resultat.

Les màquines “intel·ligents” poden efectuar les tasques que se'ls assignen cada vegada amb una eficiència més gran, però l'objectiu i el significat de les seves operacions continuaran essent determinades o habilitades per éssers humans que tenen un propi univers de valors. El risc és que els criteris que estan a la base de certes decisions es tornin menys transparents, que la responsabilitat decisional s'oculti i que els productors puguin eludir l'obligació d'actuar per al bé de la comunitat. En cert sentit, això és afavorit pel sistema tecnocràtic, que alia l'economia amb la tecnologia i privilegia el criteri de l'eficiència, tendint a ignorar tot allò que no està vinculat als seus interessos immediats¹⁰.

Això ha de fer-nos reflexionar sobre el “sentit del límit”, un aspecte sovint descuidat en la mentalitat actual, tecnocràtica i eficientista, i no obstant decisiu per al desenvolupament personal i social. L'ésser humà, en efecte, mortal per definició, pensant en sobrepassar tot límit gràcies a la tècnica, corre el risc, en l'obsessió de voler controlar-ho tot, de perdre el control de si mateix, i en la recerca d'una llibertat absoluta, de caure en l'espiral d'una dictadura tecnològica. Reconèixer i acceptar el propi límit de criatura és per a l'home condició indispensable per aconseguir o, millor, per acollir la plenitud com un do. En canvi, en el context ideològic d'un paradigma tecnocràtic, animat per una prometeica presumpció d'autosuficiència, les desigualtats podrien créixer de manera desmesurada, i el coneixement i la riquesa acumular-se en mans d'uns pocs, amb greus riscos per a les societats democràtiques i la coexistència pacífica¹¹.

5. *Temes candents per a l'ètica*

En el futur, la fiabilitat de qui demana un préstec, la idoneïtat d'un individu per a una feina, la possibilitat de reincidència d'un condemnat o el dret a rebre asil polític o assistència social podrien

⁹ Exhort. ap. *Evangelii gaudium* (24 de novembre de 2013), 233.

¹⁰ Cf. Carta. enc. *Laudato si'*, 54.

¹¹ Cf. *Discurs als participants en la Plenària de la Pontifícia Acadèmia per a la Vida* (28 de febrer de 2020).

ser determinats per sistemes d'intel·ligència artificial. La manca de nivells diversificats de mediació que aquests sistemes introdueixen està particularment exposada a formes de prejudici i discriminació. Els errors sistèmics poden multiplicar-se fàcilment, produint no només injustícies en casos concrets sinó també, per efecte dominó, autèntiques formes de desigualtat social.

A més, amb freqüència les formes d'intel·ligència artificial semblen capaces d'influir en les decisions dels individus per mitjà d'opcions predeterminades associades a estímuls i persuasions, o bé mitjançant sistemes de regulació de les opcions personals basades en l'organització de les informacions. Aquestes formes de manipulació o de control social requereixen una atenció i una supervisió acurades, i impliquen una clara responsabilitat legal per part dels productors, dels qui les usen i de les autoritats governamentals.

La dependència de processos automàtics que classifiquen els individus, per exemple, per mitjà de l'ús generalitzat de la vigilància o l'adopció de sistemes de crèdit social, també podria tenir repercussions profundes en el teixit social, establint categoritzacions impròpies entre els ciutadans. I aquests processos artificials de classificació podrien dur fins i tot a conflictes de poder, no només pel que fa a destinataris virtuals, sinó a persones de carn i ossos. El respecte fonamental per la dignitat humana postula rebutjar que la singularitat de la persona sigui identificada amb un conjunt de dades. No hem de permetre que els algoritmes determinin la manera en la que entenem els drets humans, que deixin de banda els valors essencials de la compassió, la misericòrdia i el perdó o que eliminin la possibilitat que un individu canviï i deixi enrere el passat.

En aquest context, no podem deixar de considerar l'impacte de les noves tecnologies en l'àmbit laboral. Feines que en un temps eren competència exclusiva de la mà d'obra humana són ràpidament absorbides per les aplicacions industrials de la intel·ligència artificial. També en aquest cas es corre el risc substancial d'un benefici desproporcionat per a uns pocs a costa de l'empobriment de molts. El respecte de la dignitat dels treballadors i la importància de l'ocupació per al benestar econòmic de les persones, les famílies i les societats, la seguretat de les feines i l'equitat dels salaris haurien de constituir una gran prioritat per a la comunitat internacional, a mesura que aquestes formes de tecnologia es van introduint cada vegada més en els llocs de treball.

6. Transformarem les espases en arades?

Aquests dies, mirant el món que ens envolta, no podem eludir les greus qüestions ètiques vinculades al sector dels armaments. La possibilitat de conduir operacions militars per mitjà de sistemes de control remot ha dut a una percepció menor de la devastació que aquests han causat i de la responsabilitat en el seu ús, contribuint a un apropament encara més fred i distant a la immensa tragèdia de la guerra. La recerca sobre les tecnologies emergents en el sector dels anomenats "sistemes d'armes autònoms letals", inclòs l'ús bèl·lic de la intel·ligència artificial, és un gran motiu de preocupació ètica. Els sistemes d'armes autònoms no podran ser mai subjectes moralment responsables. L'exclusiva capacitat humana de judici moral i de decisió ètica és més que un complex conjunt d'algoritmes, i aquesta capacitat no pot reduir-se a la programació d'una màquina que, tot i ser "intel·ligent", no deixa de ser sempre una màquina. Per aquest motiu, és imperiós garantir una supervisió humana adequada, significativa i coherent dels sistemes d'armes.

Tampoc no podem ignorar la possibilitat que armes sofisticades acabin en mans equivocades, facilitant, per exemple, atacs terroristes o accions dirigides a desestabilitzar institucions de govern legítimes. En resum, realment l'últim que el món necessita és que les noves tecnologies contribueixin a l'injust desenvolupament del mercat i del comerç de les armes, promovent la bogeria de la guerra. Si ho fa així, no només la intel·ligència, sinó el mateix cor de l'home correrà el risc de tornar-se cada vegada més "artificial". Les aplicacions tècniques més avançades no han d'usar-se per a facilitar la resolució violenta dels conflictes, sinó per a pavimentar els camins de la pau.

En una òptica més positiva, si la intel·ligència artificial fos utilitzada per a promoure el desenvolupament humà integral, podria introduir importants innovacions en l'agricultura, l'educació i la cultura, una millora del nivell de vida de nacions i pobles sencers, el creixement de la fraternitat humana i de l'amistat social. En definitiva, la manera en que la usem per a incloure els últims, és a dir, els germans i les germanes més febles i necessitats, és la mesura que revela la nostra humanitat.

Una mirada humana i el desig d'un futur millor per al nostre món porten a la necessitat d'un diàleg interdisciplinar destinat a un desenvolupament ètic dels algoritmes — *l'algor-ètica*—, en el que els valors orienten els itineraris de les noves tecnologies¹². Les qüestions ètiques haurien de ser tingudes en compte des de l'inici de la recerca, així com en les fases d'experimentació, planificació, distribució i comercialització. Aquest és l'enfocament de l'ètica de la planificació, en el que les institucions educatives i els responsables del procés decisonal tenen un rol essencial per exercir.

7. Reptes per a l'educació

El desenvolupament d'una tecnologia que respecti i estigui al servei de la dignitat humana té clares implicacions per a les institucions educatives i per al món de la cultura. Multiplicant les possibilitats de comunicació, les tecnologies digitals han permès noves formes de trobar-nos. No obstant això, continua essent necessària una reflexió permanent sobre el tipus de relacions al qual ens està portant. Els joves estan creixent en ambients culturals impregnats de la tecnologia i això no pot deixar de qüestionar els mètodes d'ensenyament i formació.

L'educació en l'ús de formes d'intel·ligència artificial hauria de centrar-se sobretot en promoure el pensament crític. És necessari que els usuaris de totes les edats, però sobretot els joves, desenvolupin una capacitat de discerniment en l'ús de dades i de continguts obtinguts al web o produïts per sistemes d'intel·ligència artificial. Les escoles, les universitats i les societats científiques estan cridades a ajudar els estudiants i els professionals a fer propis els aspectes socials i ètics del desenvolupament i l'ús de la tecnologia.

La formació en l'ús dels nous instruments de comunicació hauria de tenir en compte no només la desinformació, les falses notícies (*fake news*), sinó també l'inquietant augment de «pors ancestrals [...] que han sabut amagar-se i potenciar-se rere noves tecnologies»¹³. Lamentablement, una vegada més ens trobem havent de combatre «la temptació de fer una cultura de murs, d'aixecar

¹² Cf. *ibid.*

¹³ Carta enc. *Fratelli tutti* (3 d'octubre de 2020), 27.

murs, murs en el cor, murs a la terra per evitar aquest trobament amb altres cultures, amb altres persones»¹⁴ i el desenvolupament d'una coexistència pacífica i fraterna.

8. *Reptes per al desenvolupament del dret internacional*

L'abast global de la intel·ligència artificial fa evident que, juntament a la responsabilitat dels estats sobirans de disciplinar internament el seu ús, les organitzacions internacionals poden exercir un rol decisiu en la consecució d'acords multilaterals i en la coordinació de la seva aplicació i actuació¹⁵. A aquest propòsit, exhorto la comunitat de les nacions a treballar unida per a adoptar un tractat internacional vinculant, que reguli el desenvolupament i l'ús de la intel·ligència artificial en les seves múltiples formes. Naturalment, l'objectiu de la reglamentació no hauria de ser només la prevenció de les males pràctiques, sinó també encoratjar les millors pràctiques, estimulants plantejaments nous i creatius i facilitant iniciatives personals i col·lectives¹⁶.

En definitiva, en la recerca de models normatius que puguin proporcionar una guia ètica a qui desenvolupa tecnologies digitals, és indispensable identificar els valors humans que haurien d'estar a la base del compromís de les societats per a formular, adoptar i aplicar els marcs legislatius necessaris. El treball de redacció de les orientacions ètiques per a la producció de formes d'intel·ligència artificial no pot prescindir de la consideració de qüestions més profundes, relacionades amb el significat de l'existència humana, la tutela dels drets humans fonamentals, la recerca de la justícia i de la pau. Aquest procés de discerniment ètic i jurídic pot revelar-se com una valuosa ocasió per a una reflexió compartida sobre el rol que la tecnologia hauria de tenir en la nostra vida personal i comunitària i sobre com el seu ús pot contribuir a la creació d'un món més just i humà. Per aquest motiu, en els debats sobre la reglamentació de la intel·ligència artificial, s'hauria de tenir en compte la veu de totes les parts interessades, inclosos els pobres, els marginats i altres que sovint queden sense ser escoltats en els processos decisionals globals.

* * *

Espero que aquesta reflexió encoratgi a fer que els progressos en el desenvolupament de formes d'intel·ligència artificial contribueixin, en última instància, a la causa de la fraternitat humana i de la pau. No és responsabilitat d'uns pocs, sinó de tota la família humana. La pau, en efecte, és el fruit de relacions que reconeixen i acullen l'altre en la seva dignitat inalienable, i de cooperació i esforç en la recerca del desenvolupament integral de totes les persones i de tots els pobles.

La meva pregària al començament del nou any és que el ràpid desenvolupament de formes d'intel·ligència artificial no augmenti les ja nombroses desigualtats i injustícies presents al món, sinó que contribueixi a posar fi a les guerres i els conflictes, i a alleujar tantes formes de patiment que afligeixen la família humana. Que els fidels cristians, els creients de diferents religions i els homes i dones de bona voluntat puguin col·laborar en harmonia per aprofitar les oportunitats i afrontar els reptes que planteja la revolució digital, i deixar a les generacions futures un món més solidari, just i pacífic.

¹⁴ Cf. *ibid.*

¹⁵ Cf. *ibid.*, 170-175.

¹⁶ Cf. Carta enc. *Laudato si'*, 177.

Francesc

Vaticà, 8 de desembre de 2023