L'IMAGINARI SIMBÒLIC DE LA IA

La màquina (aparentment) desobedient

Kike Maíllo



Araya Peralta

Jo devia tenir uns quinze anys quan vaig veure *Blade Runner* per primera vegada.

Va ser a classe de filosofia, a l'institut. Fins aleshores, com molts nens, havia fantasiejat amb la idea de construir les meves pròpies màquines. Recordo especialment que, alguns anys abans, després de veure un capítol del *Doctor Who* on apareixien uns robots extraterrestres (els *daleks*) vaig posar la casa potes enlaire a la recerca de capses de sabates, fils de costura i oueres de cartró per fabricar els meus propis artefactes que jo mateix movia imitant aquells robots televisius. Però el que em va passar després de veure *Blade Runner* va ser diferent. Jo no aspirava a emular aquestes màquines, perquè aquestes màquines s'assemblaven massa a "nosaltres". I això m'aterria.

> Hi ha una cosa atàvica en la por que els humans sentim cap al progrés i la tecnologia. En general cap al futur que ens parla de suplantació. De substitució de la nostra raça

Hi ha una cosa atàvica en la por que els humans sentim cap al progrés i la tecnologia. En general cap al futur que ens parla de suplantació. De substitució de la nostra raça. Fruit d'aquesta por van néixer relats universals com els del *Gólem* o el *Frankestein* de Shelley, on un grapat d'insensats humans jugaven a ser déus, creant éssers als qui dotaven d"ànima".

Per a mi, el que era interessant era que a l'hora de fabricar éssers o màquines a la nostra imatge i semblança, aquests bojos i ambiciosos científics de ficció havien d'esbrinar primer "com som".

Intueixo que això mateix els passa als enginyers en robòtica o d'intel·ligència artificial en la vida real. Segurament ells també proven de trobar aquesta quadratura del cercle: la "fórmula" que defineixi les característiques essencials de la nostra manera de fer, físicament i psicològicament. A mi, particularment, l'intent d'emulació del comportament humà em resulta especialment fascinant. En aquest sentit, un dels temes centrals de *Blade Runner*, com el de molta ficció robòtica, és el de la impossibilitat del lliure albir com a paràbola sobre la pròpia llibertat humana: és possible que l'ésser humà superi les pressions del sistema? De la seva societat? De les seves emocions? I, sobretot, ens podem escapar de la química que ens compon?

Dues dècades més tard d'aquell primer visionament de *Blade Runner* i després d'atzarosos esquivaments del destí, va arribar la meva primera obra de ciència-ficció: *Eva*. La pel·lícula ens parlava d'un temps futur on proliferen les màquines socials. Robots destinats a fer companyia i combatre la solitud. Per a mi aquest assumpte, el de la solitud, era llavors un problema de rabiosa actualitat. I ho és amb més força encara ara. No és un secret que l'aïllament s'ha convertit en un problema de salut de primer ordre, especialment als països occidentals. Una epidèmia que creix de manera exponencial. Valguin unes quantes dades com a exemple [1]:

- El 25% de les llars espanyoles són habitades per un únic individu.
- Al Japó moren 4.000 ciutadans a la setmana en solitud.
- Un estudi de la Universitat de Stanford de l'any 2018 equiparava la vida en solitud a fumar 11 cigarrets al dia, que augmentava en un 31% el risc de morir, i que pot ser el detonador de malalties com la hipertensió, la demència, la depressió o els atacs de cor.

> Intueixo que la proliferació de "màquines socials" podria pal·liar, encara que fos tímidament, el problema de la solitud

Evidentment les màquines no haurien de ser l'única solució a un desafiament de tal magnitud. No tindria sentit. Però sí que intueixo que la proliferació de "màquines socials" podria pal·liar, encara que fos tímidament, el problema. I podrien fer-ho de manera extraordinàriament hàbil, sempre i quan aquestes màquines estiguessin ben dissenyades.

A a nostra pel·lícula, l'Álex, el personatge protagonista interpretat per Daniel Brühl, estava entossudit en dissenyar robots amb els quals poder compartir la vida familiar o amb els quals, per què no, anar a fer unes canyes. Ell volia dissenyar màquines que no fossin tan màquines, que no sempre es mostressin com a *esclaves* de la nostra voluntat. No deixava de ser paradoxal perquè és precisament aquesta submissió la que defineix brutalment a

qualsevol mecanisme: funcionen partint d'ordres. No oblidem que fins i tot el propi vocable "robot", prové de la paraula eslava "robota", que significa "servitud".

La dicotomia entre l'obediència i desobediència de les màquines és una constant en la ciència-ficció, però va ser expressada per primera vegada, d'una manera reglada, pel genial Isaac Asimov. Les seves famoses "tres regles de la robòtica" [2] volien limitar la conducta de les màquines intel·ligents de la següent manera:

1. Un robot no farà mal a un ésser humà o, per inacció, permetrà que un ésser humà pateixi mal.
2. Un robot ha d'obeir les ordres donades pels éssers humans, excepte si aquestes ordres entressin en conflicte amb la primera llei.
3. Un robot ha de protegir la seva pròpia existència en la mesura que aquesta protecció no entri en conflicte amb la primera o la segona llei.

La idea d'aquestes tres lleis de ficció era protegir l'ésser humà per tal que els robots, per molt avançats que fossin, no poguessin girar-se en contra nostra, com succeïa amb el *Frankestein*, de Mary Shelley.

Però en el camp de la màquina social o de companyia, aquella que ha d'ocupar una esfera estrictament personal, s'obre un debat molt interessant: podran ser bones companyes de viatge aquelles màquines que només i absolutament només obeeixin sempre les ordres del seu "amo"? Podem establir relacions poderoses amb individus que no sentim com a "iguals"?

Es tracta d'un debat interessant. Encara que no sóc més que un aficionat, sé que són diverses les investigacions que avancen emulant l'afectivitat humana i la comprensió de "l'altre" basant-se en la Intel·ligència Artificial. Sembla bastant raonable que les màquines del futur sàpiguen llegir els nostres estats d'ànim i comportar-se de forma compassiva, atenta i tendra si volem que siguin bons companys de viatge.

Però estic segur que arribarem més enllà.

No em costa d'imaginar que en el futur serem capaços de dissenyar robots amb els quals establirem una sincera amistat.

> La possibilitat que una màquina pugui no obeir-nos i fins i tot rebutjar-nos em sembla crucial si volem "dissenyar" màquines que "emulin" el lliure albir i amb les quals, per tant, puguem crear ponts emocionals a un nivell superior

Existeix una apreciació universalment acceptada que l'amistat veritable es basa no només

en l'afectivitat i empatia que existeix entre dos o més persones, sinó principalment en el respecte mutu. La meva limitada experiència personal i intel·lectual m'ha ensenyat que, malgrat que la majoria de nosaltres establim diferents tipus de relacions personals, aquelles que són més fortes (i aquest seria l'objectiu ulterior d'una màquina social o de companyia) s'estableixen amb aquells individus que sentim com a "iguals". Entenem que estem davant d'un "igual" quan aquest individu és autònom i com a tal pot exercir aquesta independència en una direcció diferent o fins i tot oposada a la nostra. Potser això es deu a que els humans, com a espècie, som animals gregaris i, de la mateixa manera que els passa a molts altres mamífers, ens sentim reconfortats en seguir un guia o recolzar una consigna exposada per un grup de lideratge. Per aquesta raó, quan algú no segueix les ordres, d'alguna manera "se significa", sobresurt. En anar a contracorrent, adquireix "valor". La possibilitat que una màquina pugui no obeir-nos i fins i tot rebutjar-nos em sembla crucial si volem "dissenyar" màquines que "emulin" el lliure albir i amb les quals, per tant, puguem crear ponts emocionals a un nivell superior.

No descobreixo res d'extraordinari si afirmo que moltes de les nostres experiències més fortes giren entorn del desenllaç positiu després d'una crisi. A la catarsi que hem sentit en solucionar un conflicte. Diverses estratègies pedagògiques o narratives es basen en aquest mateix dispositiu: es crea un problema que acabem o no resolent. Especialment la narrativa universal (en les seves diferents variants: literatura, cinema o videojocs) sol operar proposant-nos una meta, un objectiu que ha de ser assolit. Després, l'esdevenir de la ficció no deixa de crear obstacles i traves per aconseguir-ho, allunyant aquesta meta i fent més emocionant el viatge fins a arribar a un clímax final.

De la mateixa manera, crec que aquestes màquines socials del futur haurien de ser mestres del conflicte, del noble art de la narrativa aplicada a les relacions. Naturalment haurien de ser molt hàbils en saber com i quan crear un obstacle. Gairebé ningú no vol tenir com a company algú que sigui un mur, que ens porti la contrària per esport. Es tractaria, d'alguna manera, que la segona llei de la robòtica d'Asimov, la que ens parla de l'obligació d'obediència, fos una mica més difusa, menys categòrica, més creativa. Es tractaria d'alliberar la màquina de la submissió als nostres capricis.

> El robot de companyia del futur que imagino hauria de ser, per tant, un individu molt adaptable a la "llar" on arribi. Per això, aquesta intel·ligència artificial hauria de ser hàbil en psicografia, en l'estudi i classificació de persones segons el seu caràcter

El robot de companyia del futur que imagino hauria de ser, per tant, un individu molt adaptable a la "llar" on arribi. Jo tinc un nivell d'acceptació de la formalitat, l'afecte o el conflicte en les relacions diferent al que, probablement, tingui el meu veí. És més, el meu veí és incapaç de mantenir el seu humor constant durant tota la setmana. Per això, aquesta intel·ligència artificial hauria de ser hàbil en psicografia, en l'estudi i classificació de persones segons el seu caràcter. En l'actualitat són diversos els sistemes d'aquest tipus de

categorització de la personalitat, alguns d'ells ja implementats en la Intel·ligència artificial. Des del model dels Cinc Grans (OCEAN: *Openness* / *Conscientiousnes* / *Extraversion* / *Agreeableness* / *Neuroticism*) [3], a l'estructura jeràrquica de la personalitat d'Eysenk (PEN: el tret de caràcter pot explicar-se partint del Psicoticisme, Neuroticisme i Extraversió) [4], passant per la teoria de la personalitat de Cattell (basat en 16 trets de personalitat o factors font) [5]. Qualsevol d'aquests sistemes podria ajudar a aquests robots a crear-se una idea del nostre caràcter i a elaborar un itinerari o full de ruta de relació personal profunda, inspiradora i, ja que hi som, entretinguda. Adaptant-se a nosaltres a cada pas.

Per descomptat aquestes idees són tan sols conjectures d'algú que sol moure's en el marc de la ciència-ficció i, per tant, en un laboratori de proves gairebé sempre favorable. La realitat científica és un assumpte diferent. O potser no tant... En qualsevol cas, aquests són els apunts d'un *xaval* que va començar interessant-se en l'aspecte dels robots per acabar captivat pel comportament d'aquests. Que intueix que una de les pròximes revolucions, que no l'única, la trobarem a les màquines que són capaces de ser desobedients sense deixar de ser divertides, afectuoses i, per descomptat, segures.

REFERÈNCIES

1 — El País. Negocis. 12/24/2018. "La factura econòmica de la solitud".
2 — Les normes van aparèixer per primer cop al conte curt de 1942 *Runaround*, que forma part de la col·lecció de 1950 *Jo, robot*.
3 — Psychological Review, 41, 1-32.
4 — Psychologistworld.com "Eysenck's PEN Model of Personality"
5 — Verywellmind.com "Cattell's 16 Personality Factors".



Kike Maíllo
Kike Maíllo és graduat en direcció per l'ESCAC (Escola Superior de Cine i Audiovisuals de Catalunya). Ha treballat com a creatiu i realitzador, i ha fet nombrosos videoclips d'èxit. La seva carrera va iniciar-se amb el seu segon curtmetratge, "Los perros de Pavlov", que va ser projectat en més de quaranta festivals i va rebre més de dues desenes de premis, entre els que destaquen els del Festival de Alcalá de Henares o el Festival de Munich. L'any 2009 s'encarrega de la posada en marxa i la realització de la sèrie d'animació "Arròs covat", produïda per Escándalo Films i Televisió de Catalunya, i guanyadora d'un premi Ondas l'any 2010. A l'octubre del 2011 estrena "Eva", el seu primer llargmetratge, un drama de ciència ficció que ha rebut més d'una vintena de premis, entre els quals destaca el Goya al millor director novell. El 2016 estrena "Toro", el seu segon llargmetratge, i a l'actualitat prepara el seu tercer film com a productor, "La llamada".