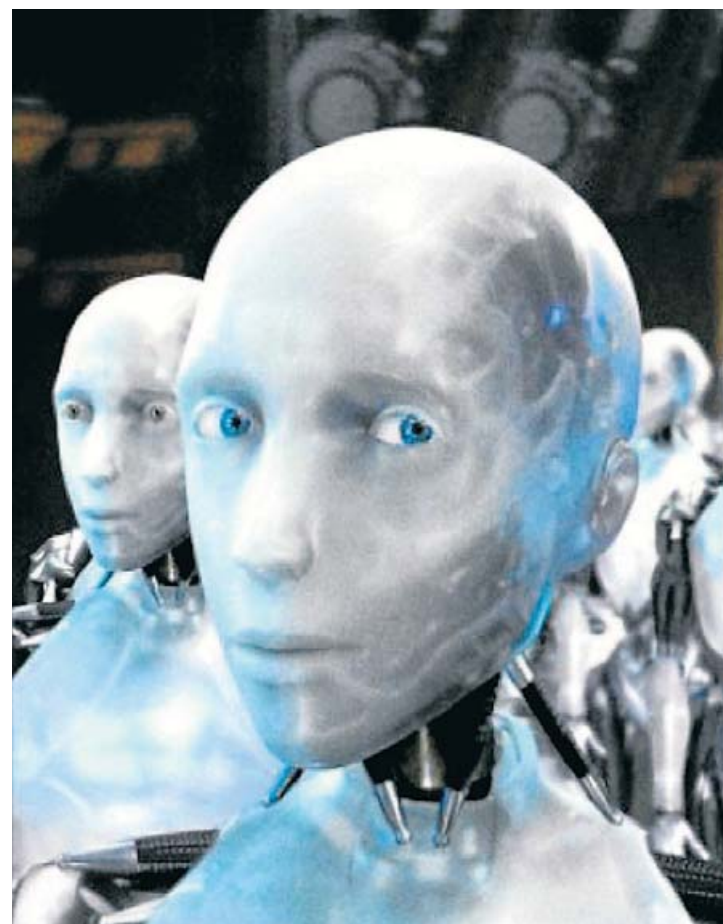


LES LLEIS D'ASIMOV El nucli de la pel·lícula *Jo, robot* (2004), dirigida per Alex Proyas, qüestiona les tres lleis de la robòtica d'Isaac Asimov. Es resumeixen en la llei zero: un robot no pot fer mal a la humanitat.

20th Century Fox



D'aquí tres anys hi haurà deu milions de cotxes sense conductor circulant per les carreteres, al mercat ja hi ha robots dedicats a cuidar persones i els empresaris es preparen per introduir-los en massa en substitució del treballadors humans. No són escenaris de ciència-ficció, ja està passant. Fins a quin punt s'ha de regular legalment amb quina ètica es programen els robots? Tindran drets?

LA MORAL DELS ROBOTS

TEXT

Xavier **Vidal** / Gina **Tost**



Fa uns mesos, Microsoft va haver de tancar el perfil de Tay, la primera intel·ligència artificial a tenir un compte de Twitter. Poques hores després de començar a piular, Tay es va tornar racista, sexista i, fins i tot, pronazi en interactuar amb altres tuitaires. ¿Els robots poden desenvolupar una moral o sentiments propis que guiïn els seus actes més enllà dels algorismes amb què han estat programats? Ningú ho pot dir, de moment, però el que sembla evident entre els especialistes és que cal dissenyar unes normes adequades que regeixin les relacions entre les persones i les noves màquines connectades, com ara els cotxes autònoms o els robots domèstics. És a dir, actualitzar les lleis de la robòtica que Isaac Asimov va formular des de l'àmbit de la ficció fa més de setanta anys.

La regulació dels comportaments dels robots no només és una preocupació industrial sinó també política. La comissió d'assumptes jurídics del Parlament Europeu està treballant en un informe sobre el dret civil de la robòtica. En el document inicial s'alerta que en un futur proper la intel·ligència artificial (AI) podria superar la capacitat intel·lectual humana i, si no s'està preparat, això podria suposar un doble desafiament per a la humanitat: "La capacitat de ser amos del seu propi destí [el dels robots] i garantir la

supervivència de l'espècie". El document cita com a base les lleis formulades per Isaac Asimov i planteja una qüestió clau: si l'autonomia dels robots hauria d'implicar la creació d'una nova categoria civil més enllà de persones físiques, jurídiques, animals o objectes per delimitar "l'atribució de drets i obligacions" dels nous androïdes amb capacitat d'operar de manera autònoma. Isaac Asimov i John W. Campbell van desenvolupar tres lleis de la robòtica per al conte *Runaround* [El cercle viciós] l'any 1942. Les lleis diuen que un robot no pot fer mal a un ésser humà o, per inacció, permetre que un ésser humà prengui mal; que un robot ha d'obeir les ordres dels éssers humans, excepte si entren en conflicte amb la primera llei, i que 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.

Anys més tard se'n va formular una altra, la llei zero, que diu que un robot no pot fer mal a la humanitat o, per inacció, permetre que la humanitat prengui cap mal. Les lleis d'Asimov van ser pensades per protegir les persones i evitar el "complex de Frankenstein", és a dir, que les màquines assolissin un nivell de desenvolupa-

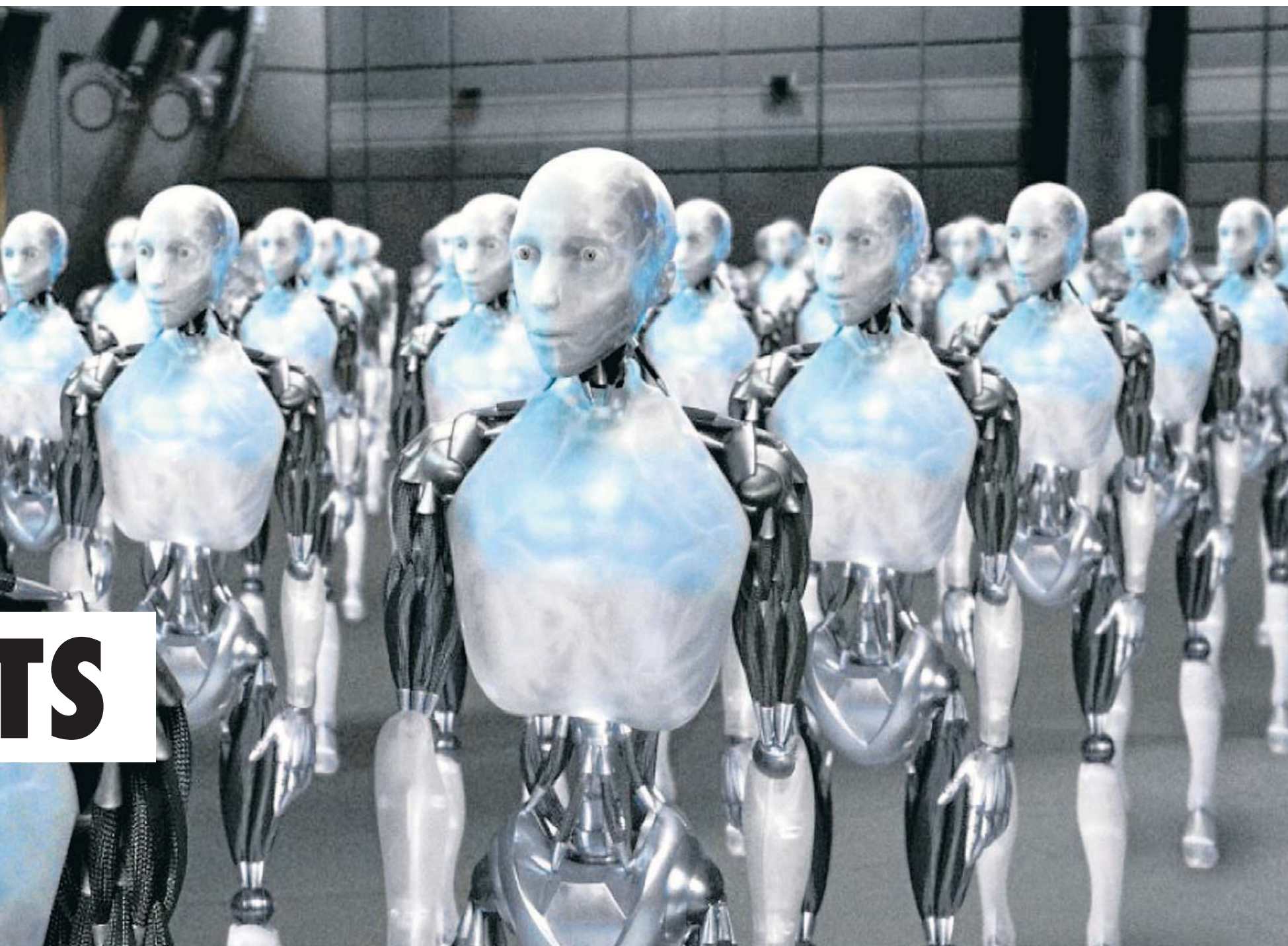
ment tal que poguessin rebel·lar-se contra els seus creadors, la hipòtesi inicial a la pel·lícula *Terminator*.

Moral robòtica?

Per a Carme Torras, professora d'investigació del CSIC a l'Institut de Robòtica i Informàtica Industrial, la llei zero és la més important i on cal posar l'accent és en la formació de "les persones que dissenyen i programen robots, així com en les que tenen relació amb ells". Torras afirma que "s'està intentant programar una moral, per exemple, en els cotxes autònoms i establir uns principis bàsics en l'internet de les coses". Els cotxes autònoms són la preocupació més immediata. Sembla llunyà haver-se d'enfrontar als perills que plantejava la Skynet revoltada contra els humans a *Terminator* o la rebel·lió dels Ns-5 de *Jo, robot*, però els cotxes sense

conductor ja són una realitat. Segon uns informe de Business Insider, d'aquí tres anys hi haurà deu milions de cotxes sense conductor circulant per les carreteres. Fa uns mesos, un grup d'estudi de l'MIT va publicar un document anomenat *El dilema social dels vehicles autònoms*. Plantejava una hipòtesi molt real: ¿a qui triaria sacrificar un

• *El Parlament Europeu està treballant sobre el dret civil de la robòtica* •



TS

cotxe autònom en cas d'haver d'escollir: al seu passatger o a un grup, per exemple, de vianants o ciclistes? En bona lògica el vehicle hauria de salvar el màxim nombre d'humans però, aleshores, ¿qui comprarà un cotxe que, sota certes circumstàncies, pot decidir que mori el seu propietari? És una equació per a la qual encara no hi ha una resposta clara.

Aquest exemple fa pensar directament en la moral i l'ètica i no només en les regles estrictes i objectives que hauria de seguir un robot autònom en les situacions més compromeses. Raúl Rojas, professor d'intel·ligència artificial a la Universitat Lliure de Berlín, creu que s'ha d'anar més enllà d'uns protocols tancats perquè les persones "de vegades hem de sortir de les regles per solucionar un problema, identificar-ne la contradicció i pensar a un altre nivell". I posa un exemple: si un robot està programat per respectar els semàfors i un semàfor s'espatlla i es queda en vermell, el robot no passarà mai perquè està programat per respectar l'ordre de pas, mentre que un conductor sabrà que no funciona i se'l saltarà per superar aquella situació.

No tots els casos són tan nítids com el del semàfor. Els robots dedicats a cuidar la nostra salut ja

són al mercat i aviat formaran part de la nostra quotidianitat. En el moment en què es facin autònoms actuaran de manera independent sobre els nostres tractaments vitals. Miquel Montaner, assessor tecnològic independent a Cumlaude.tech, creu que una màquina "no hauria de poder prendre decisions sobre aspectes que no hagin sigut prèviament discutits i acceptats pels experts".

• *"Si un robot m'ha d'operar el cor, voldria que abans m'ho consultés"* •

Raúl Rojas opina que en situacions de dubte ètic o moral el robot hauria de consultar un ésser humà: "Si una computadora decideix operar-me el cor, m'agradaria que abans m'ho consultés a mi o a un metge de carn i ossos". Però preguntar a una persona què s'ha de fer en determinades situacions obre un altre dubte: ¿i si l'humà ordena a la màquina que se salti les normes?

En el cas dels cotxes, per exemple, Raúl Rojas, que va adaptar les lleis d'Asimov als vehicles autònoms, ho té clar: el cotxe ha de respectar les normes de trànsit per sobre del que li demani el propietari. Aleshores, però, cal preguntar-se si s'hauria de permetre a un robot saltar-se les normes a fi d'evitar un mal més gran. I aquesta prerrogativa entra de ple en el terreny de l'ètica i la valoració subjectiva. És a dir, que cal donar llibertat de decisió al robot.

Sembla que no tot és objectivable en les relacions entre robots i persones i que hi ha també una part emocional que cal tenir en compte, tot i que més aviat dels humans cap a les màquines que no pas al revés.

Vincles sentimentals

El projecte del Parlament Europeu sobre la robòtica afirma que en el cas dels robots assistencials, les persones "podrien desenvolupar sentiments emocionals", la qual cosa planteja la "preocupació per la dignitat humana i altres valors morals". Carme Torras no creu que "els robots arribin a tenir sentiments com nosaltres", és a dir, entesos com a resultat de les nostres experiències. Ara bé, "una altra cosa és programar-los perquè facin veure que en tenen", assegura. Miquel Montaner explica que s'està treballant en la "simulació computacional de sentiments, tot i que no són sentiments propis, sinó preprogramats perquè les persones percebin que ho són". I això planteja el debat sobre si és ètic que les màquines simulin sentiments. Carme Torras diu que un dels principis bàsics en ètica és no enganyar les persones i "un robot hominoide pot enganyar la persona que cuida i aquesta persona pot arribar a creure que el robot realment es preocupa pel seu benestar i delegar-li totes les decisions". O, en el cas d'un nen, que pugui arribar a pensar que el seu assistent robòtic és un veritable amic. "Només



LES LLÀGRIMES DEL REPLICANT El replicant Roy Batty (Rutger Hauer) plora per la seva desaparició i l'egoisme dels humans en un dels discursos més emocionants de la ciència-ficció al final de *Blade runner* (1982), un film basat en el llibre *Els andròides somien xaïls elèctrics?*, de Philip K. Dick. Warner Bros

em sembla acceptable si és un adult qui, plenament conscient de les capacitats d'una màquina, decideix tenir-hi relacions sentimentals, com a la pel·lícula *Her*, posa com a exemple Carme Torras. En aquesta pel·lícula de Spike Jonze, el protagonista s'enamora del sistema operatiu femení que li gestiona totes les tasques vitals.

Màquines o esclaus

Els sentiments i el sexe entre robots i humans són hipòtesis molt explotades en la literatura i el cinema. A la pel·lícula *Blade runner* (1981), basada en l'obra del mestre de la ciència-ficció Philip K. Dick, l'empresa Tyrell Corporation fabrica un tipus d'androide femení només destinat a satisfer els desitjos sexuals dels homes. Pot semblar pura ficció però a Londres està a punt d'obrir un bar on s'oferiran serveis de fel·lació fets per robots sexuals femenins a seixanta lliures, cafè inclòs. Potser ara mateix aquestes esclaves sexuals només són simples màquines, però ¿què passarà d'aquí uns anys, quan evolucionin i s'assemblin físicament a una dona de carn i ossos i tinguin una intel·ligència artificial que els doti de sentiments simulats o empatia? ¿Arribarà un punt en què hauran de tenir drets, encara que no siguin com els de les persones? ¿Fins a quin punt se'ls ha de respectar?

Un de les sèries més exitoses aquesta temporada ha sigut *Westworld*, basada en la novel·la i la pel·lícula creada per un dels escriptors de culte de la ciència-ficció, Michael Crichton, l'any 1973. La producció narra la vida en un parc temàtic de robots amb comportament i aparença totalment humans i on les persones poden anar a matar, violar i abusar dels robots amb impunitat, emparant-se justament en la seva condició de màquines.

Miquel Montaner no té dubtes quan diu que “els drets són patrimoni dels éssers vius, no de les màquines, encara que les hàgim desenvolupat perquè semblin humans”. La semblança física amb els humans, per descomptat, hi fa molt a l'hora de pensar en el possible tracte que

hem de tenir amb els andròides. Carme Torras opina que “ara per ara els robots són eines i no cal fer-los més humans del que és necessari”. En la seva novel·la *La mutació sentimental*, un dels personatges diu que “els robots malcriadors fan persones malcriades, els robots esclaus fan dèspotes i els robots entretenidors renten el cervell als seus propietaris”. Sembla clar, per tant, que algun tipus de respecte i tracte haurem de tenir amb les nostres màquines a mesura que s'assemblin més a nosaltres mateixos.

Beneficis contra perills

De moment, però, aquest terreny sembla només patrimoni de la ficció, i la hipòtesi de considerar les màquines com a esclaus no és aplicable en el nivell de desenvolupament robòtic actual. Raúl Rojas afirma que “totes les màquines són els nostres esclaus, començant per la rentadora o l'aspiradora, però com que no tenen voluntat pròpia ni sentiments” no se'ls podria aplicar el concepte d'esclavatge com l'entendem amb les persones. S'ha d'anar amb compte, però. L'informe del Parlament Europeu planteja un escenari en què els robots facin gran part de la feina que ara mateix fan els éssers humans i això “genera interrogants sobre el futur de l'ocupació i la viabilitat dels sistemes de seguretat social [...] que podrien comportar una desigualtat més gran en la distribució de la riquesa i el poder” en la societat.

Carme Torras té clar que “els empresaris que utilitzin massivament robots en comptes de contractar persones potser sí que hauran de cotitzar per ells a la seguretat social”.

Potser un dels elements bàsics per determinar com han de ser les relacions entre humans i màquines en un futur immediat és definir quin tipus de robots volem com a societat. Per exemple, si hem de tenir robots disse-

nyats per matar en la indústria militar. Fa uns mesos, la policia de Dallas va enviar un robot bomba per matar Micah Xavier Johnson, un dels sospitosos del tiroteig de Dallas. L'androide va fer efectives les ordres de la policia i va abatre el tirador. Quines lleis de comportament se li han de donar a un robot que té com a finalitat matar humans?

Carme Torras explica que recentment s'ha elaborat un manifest contra els *killer robots* que ella mateixa ha firmat. Raúl Rojas hi està d'acord i creu que aquest tipus de màquines “no s'han de construir ni vendre i s'ha de fer una convenció internacional que les prohibeixi”. Miquel Montaner té una altra visió quan explica que s'ha de diferenciar entre robots i eines robòtiques. “En el cas de Dallas –diu Montaner– la decisió de matar va ser d'una persona i no d'un robot”.

En el punt neuràlgic de tot el debat hi ha el dubte de si els robots poden arribar a desen-

volupar algun tipus d'autoconsciència o alguna idea propera al que els humans considerem l'ànima. Potser si es considera que no tenen ànima queda justificat que no tinguin drets o obligacions en el sentit humà del terme. És difícil predir si això passarà, però no fa tant de temps que *teòrics* del comportament social, com ara Francis Galton a principis del segle XIX, sostenien que els negres o les dones eren éssers inferiors i basant-se en això se'ls negaven els mateixos drets que als homes blancs. I

no era precisament un autor poc llegit o de teories estrambòtiques; era el cosí de Darwin i va tenir influència en la formulació de la seva teoria de l'evolució.

I això ens porta directament a la sèrie de la HBO *Westworld*: sabent que són màquines però amb aspecte i reaccions semblants a les humans, ¿ens podem permetre tractar de manera horrible o sense ètica els nostres robots? ●

•
¿Ens podem
permetre
tractar sense
ètica els
nostres
robots?
•