
Intel·ligència artificial i robòtica

PID_00270381

Pablo Jiménez Schlegl
Carme Torras Genís

Temps mínim de dedicació recomanat: 6 hores



**Pablo Jiménez Schlegl**

Pablo Jiménez Schlegl és doctor enginyer industrial per l'UPC i científic titular del CSIC a l'Institut de Robòtica i Informàtica Industrial (CSIC-UPC), on actualment ocupa el càrrec de cap de departament. Els seus interessos científics inclouen la planificació de tasques, l'aprenentatge, i més recentment la manipulació robòtica d'objectes deformables d'ús diari, com la roba, amb l'objectiu de contribuir a la robòtica assistencial. Darrerament també s'ha interessat per temes relacionats amb l'impacte de la robòtica en la societat humana des d'un punt de vista ètic.

**Carme Torras Genís**

Carme Torras és llicenciada en matemàtiques, doctora en informàtica i professora de recerca en l'Institut de Robòtica e Informàtica Industrial (CSIC-UPC) en Barcelona, on dirigeix un grup d'investigació en robòtica assistencial.

Es *Fellow* del IEEE i l'EurAI, membre numerari de la Reial Acadèmia de les Ciències i Arts de Barcelona, i de l'Acadèmia Europea; i ha obtingut reconeixements com la medalla Narcís Monturiol de la Generalitat de Catalunya i el premi Julio Peláez a les dones pioneres en ciències físiques, químiques i matemàtiques. La seva novel·la de ciència-ficció, *La mutación sentimental*, ha estat publicada en anglès amb el títol *The Vestigial Heart*.

L'encàrrec i la creació d'aquest recurs d'aprenentatge UOC han estat coordinats per la professora: Mònica Vilasau Solana (2020)

Primera edició: febrer 2020
© Pablo Jiménez Schlegl, Carme Torras Genís
Tots els drets reservats
© d'aquesta edició, FUOC, 2020
Av. Tibidabo, 39-43, 08035 Barcelona
Realització editorial: FUOC

Cap part d'aquesta publicació, incloent-hi el disseny general i la coberta, no pot ser copiada, reproduïda, emmagatzemada o transmesa de cap manera ni per cap mitjà, tant si és elèctric com químic, mecànic, òptic, de gravació, de fotocòpia o per altres mètodes, sense l'autorització prèvia per escrit dels titulars dels drets.

Índex

1. Robòtica i intel·ligència artificial per a profans.....	5
1.1. Robots i intel·ligència artificial	5
1.2. Què és un robot?	7
1.3. Aplicacions	12
1.4. Tècniques d'intel·ligència artificial en robòtica	17
2. Consideracions ètiques.....	21
2.1. És possible construir màquines ètiques?	22
2.2. Rellevància ètica dels robots en diferents àmbits de l'experiència humana	27
2.2.1. L'àmbit laboral: impacte del robot sobre el mercat de treball	28
2.2.2. L'àmbit del transport: vehicles autònoms	31
2.2.3. L'àmbit del conflicte: robots militars	33
2.2.4. L'àmbit de la salut: pròtesis, sistemes experts i robots cirurgians	37
2.2.5. L'àmbit domèstic: seguretat, privadesa, respecte a la intimitat	40
2.2.6. L'àmbit de la intimitat: relacions afectives	42
3. Iniciatives legals.....	47
3.1. La intel·ligència artificial i el món de les lleis	47
3.2. Responsabilitat civil i penal dels actors implicats en el desenvolupament del robot intel·ligent	49
3.2.1. Accidents: els robots poden causar dany	49
3.2.2. Un marc legal per a robots	51
3.3. Iniciatives legals en l'àmbit europeu	57
3.3.1. Precedents	57
3.3.2. Iniciativa legislativa del Parlament Europeu	58
3.4. Iniciatives normatives i legals en l'àmbit internacional i extracomunitari	60
3.4.1. Normes ISO/TC 299 Robotics	60
3.4.2. Àmbit extracomunitari	62
3.5. Drets del robot: visions de futur	63
Resum.....	66
Bibliografia.....	69

1. Robòtica i intel·ligència artificial per a profans

No és necessari ser un expert tecnòleg ni tan sols un aficionat a la ciència-ficció perquè els termes intel·ligència artificial i robot tinguin moltes connotacions familiars. També sabem que comencen a exercir un paper cada vegada més rellevant en les nostres vides. Sabem que en molts aspectes ens faran la vida més fàcil (aquest és el seu objectiu) però que també seran font i objecte de conflictes, disputes, accidents, vulnerabilitats i altres efectes negatius. La veritat és que han vingut per quedar-se, i dels humans depèn que el seu desenvolupament i expansió sigui no solament innocu sinó beneficiós en conjunt. Per això és decisiu que en primer lloc sapiguem de què parlem realment. Aquest és l'objectiu d'aquesta primera part: assentar les bases terminològiques i conceptuals.

1.1. Robots i intel·ligència artificial

La **intel·ligència artificial** (IA) comprèn tot el conjunt de teories, mètodes i algorismes que pretenen emular processos cognitius. Encara que hi ha altres tipus d'intel·ligència al planeta (que revesteix formes variades, com la intel·ligència mostrada per mamífers o la intel·ligència col·lectiva d'eixams i formiguers), el referent principal de l'IA és la intel·ligència humana. En un llibre divulgatiu (López de Mántaras i Meseguer, 2017), els experts en IA López de Mántaras i Meseguer plantegen la doble pregunta de si és possible construir màquines intel·ligents i si el cervell és una màquina. Tal com afirmen aquests autors, «l'objectiu últim de l' IA –aconseguir que una màquina tingui una intel·ligència *general* similar a la humana– és un dels més ambiciosos que s'ha plantejat la ciència» (López de Mántaras i Meseguer, 2017, pàg. 7). L'èmfasi en la citació original dona motiu per a diferenciar l'IA *específica* de la *general*, la primera de les quals és la inherent als sistemes experts (SE) i els programes capaços de jugar a escacs (que arriben a derrotar campions mundials) o a pòquer, mentre que la segona ha d'exhibir la versatilitat de la intel·ligència humana. Una altra distinció destacada igualment és entre IA *feble* i IA *forta*, la primera de les quals és sinònima d'específica, mentre que la segona, a més de ser general (la reciprocitat no és certa, és a dir, no són sinònimes), va més enllà de reproduir o simular els processos mentals: *és una ment*. Alguns autors, com el filòsof John Searle, qüestionen que es pugui construir una IA forta, que il·lustra amb l'experiment mental de l'habitació xinesa: una persona equipada amb un manual anglès-xinès en una habitació pot proporcionar traduccions a l'anglès dels ideogrames que rep, però això no significa que sap realment xinès (és a dir, es limita a establir correspondències). Al marge d'aquesta discussió, la veritat és que l'IA feble ha proporcionat molts èxits i aplicacions de gran

utilitat, com els sistemes experts ja esmentats, i el que pot fer de manera molt eficient és, en lloc de duplicar l'activitat mental humana, complementar-la gràcies a les seves capacitats de processament i de gestió de la memòria.

Seguint amb el text de López de Mántaras i Meseguer, hi ha quatre tipus de models d'IA:

- **Simbòlic.** Es basa en la hipòtesi dels sistemes de símbols físics (SSF), que afirma que «tot sistema de símbols físics té els mitjans necessaris i suficients per a dur a terme accions intel·ligents». Aquest ha estat el model dominant en IA i és «un model *top-down* que es basa en el raonament lògic i la cerca heurística com a pilars per a resoldre problemes» (López de Mántaras i Meseguer, 2017, pàg. 11). La majoria dels sistemes experts als quals hem al·ludit abans estan basats en aquest model: els símbols que manegen són paraules de la terminologia característica del domini especialitzat en qüestió (metge, legal, enginyer, etc.).
- **Connexionista.** És un model *bottom-up* que considera que l'activitat intel·ligent emergeix a partir de processos més simples molt interconnectats. La versió més estesa és la de la xarxa neuronal artificial, en la qual els processos simples es modelitzen com a neurones, les interconnexions corresponen a les sinapsis i els processos d'entrenament de la xarxa (modificació local de la intensitat de les connexions) emulen l'aprenentatge de les correlacions entre entrades i sortides del sistema.
- **Evolutiu.** En aquest model els comportaments intel·ligents emergeixen en un sistema que emula els mecanismes de l'evolució biològica: la informació continguda en cromosomes està subjecta a mutacions aleatòries, a les lleis de l'herència genètica i de la selecció natural. Els individus que aconsegueixen sobreviure i prosperar representen processos mentals millor adaptats a la resolució dels problemes plantejats per un entorn específic (que és la categoria de problemes a la resolució dels quals va dirigit el sistema).
- **Corpori.** En aquest model es posa èmfasi en la *cognició situada*: per a dotar els mecanismes cognitius de contingut semàntic, és necessari que experimentin el seu entorn directament mitjançant sensors i puguin també modificar-lo amb les seves accions. Tal com diuen López de Mántaras i Meseguer, molts investigadors (entre els quals hi ha ells mateixos) estan convençuts que l'aproximació corpòria és «imprescindible per a avançar cap a intel·ligències de tipus general» i que «el cos dona forma a la intel·ligència» (López de Mántaras i Meseguer, 2017, pàg. 14).

Allen Newell i Herbert Simon

López de Mántaras i Meseguer citen Allen Newell i Herbert Simon en la seva ponència en rebre el premi Turing de 1975.

Reflexió

Encara que el model connexionista exhibeix paral·lelismes evidents amb el funcionament del cervell, encara és molt lluny de capturar-ne la immensa complexitat. Sobre aquest tema, vegeu López de Mántaras i Meseguer (2017, pàg. 13).

La intel·ligència humana és òbviament una font d'inspiració de tots aquests models, però així com el primer és un intent de formulació abstracta dels processos mentals, els altres tres pretenen emular els mecanismes físics que han portat a aquesta formulació. De tots aquests models, l'últim és el que ens in-

teressa més en el present curs. Cal subratllar que no és incompatible amb els anteriors; de fet, pot recórrer-hi com a model de representació interna. El cos en qüestió és artificial com la intel·ligència que «encarna». Per descomptat, parlem del robot.

En definitiva, un robot dona cos a l'IA, li permet percebre el món de primera mà, moure's per l'entorn físic, actuar. Recíprocament, gràcies a les tècniques de l'IA, el robot no és una màquina i prou: és una màquina capaç de recordar, aprendre, planificar, raonar, prendre decisions més o menys complexes. Per descomptat, l'IA incorporària té moltes aplicacions i implicacions ètiques i legals pels danys que pot ocasionar potencialment. Al llarg d'aquest mòdul es comentaran els casos més rellevants, però l'èmfasi es posarà en l'IA que actua al món físic mitjançant la robòtica. Per a això, en primer lloc hem d'aclarir què és un robot.

1.2. Què és un robot?

Tothom creu que sap què és un robot. Aquesta paraula, encunyada fa un segle per l'escriptor txec Karel Čapek en la seva obra del 1920 *R.U.R. (Rossum's Universal Robots)*, evoca preferentment un humanoide artificial, metàl·lic, que fa moviments bruscos, sincopats i té veu enllaunada. La ficció televisiva i cinematogràfica ha anat ampliant el rang d'accepcions d'aquesta imatge, per la qual cosa l'imaginari popular no té inconvenient a admetre en la mateixa categoria el símil de paperera o aspiradora rodant que és R2D2 de *Star Wars*, i els replicants pràcticament indistingibles dels humans de *Blade Runner*.

Però molt poques persones saben què és un robot realment. Un robot ha de tenir forma humana, si més no vagament? Un autòmat del segle XVII és un robot? Si la forma humana no és rellevant, pot un vehicle autònom ser considerat un robot? I una rentadora intel·ligent? Totes les màquines que tenen capacitat de processar informació externa són robots? Un PC és un robot? Un dels pares de la robòtica industrial (una categoria de robots ben definida), Joseph Engelberger, va declarar en una ocasió que ell no sabia definir un «robot», però que sabia reconèixer-ne un quan el veia. No hi ha un consens generalitzat sobre la definició precisa de robot, i hi ha moltes discussions sobre si una determinada màquina es pot qualificar com a tal, però sí que hi ha un cert acord en què el requisit fonamental es pot resumir en **el cicle percep-pensa/planifica-actua:**

Rossum's Universal Robots

Fou estrenada el 25 de gener de 1921 al Teatre Nacional de Praga. El terme utilitzat per a designar els humanoides fabricats per a treballar en lloc dels humans va ser suggerit al dramaturg pel seu germà pintor Josef.

- Un robot ha de ser capaç de percebre el seu entorn mitjançant sensors (depenent del que s'espera de la màquina, aquests sensors seran més o menys sofisticats).
- Si hi ha diferents alternatives, ha de ser capaç de decidir quin és el curs d'acció més convenient per aconseguir els objectius encomanats d'acord amb la seva programació (el terme *pensar* incomoda molts científics perquè designa una activitat humana sobre la qual encara hi ha molts interrogants, i ells prefereixen els termes més precisos de *presa de decisions* per avaluar accions d'efectes immediats o *planificació* si es tracta de determinar una seqüència d'accions que portaran a la meta).
- Ha de ser capaç de dur a terme accions en un entorn físic mitjançant els seus actuadors, accions que van des de mers desplaçaments fins al trasllat o manipulació d'objectes, modificació de la seva pròpia estructura, interacció amb persones, etc.

Segons la norma ISO 8373:2012, un robot és:

«Un mecanisme actuat programable en dos o més eixos amb un grau d'autonomia, que es mou en el seu entorn per fer les tasques previstes.»

Així, doncs, ningú no qualificaria de robot un termòstat, però tampoc una casa domòtica on un ordinador central, programat d'acord amb les preferències dels seus habitants, regula el conjunt de termòstats de la llar, a més dels llums, persianes, neteja, comunicacions o accessos: per més sofisticada que sigui, i malgrat ser programable, falta la component del *moviment* en un entorn (aquest argument no és incontrovertible, ja que l'obertura i tancament de finestres comporta el moviment d'una part del conjunt, com en un robot reconfigurable)¹. És més subtil el cas d'un automatisme d'actuadors pneumàtics (sistemes cilindre-èmbol actuats per aire a pressió): pot ser programat si és controlat mitjançant un PLC, però aquesta programació es refereix solament a la *seqüència*, l'ordre en què són activats els actuadors, però no es pot variar la seva *posició* per programa, la qual ve donada pels extrems del recorregut de l'èmbol. Per tant, tampoc no són robots, ja que aquests poden posicionar-se en qualsevol punt del seu espai de treball.

D'acord amb el seu aspecte físic, molt lligat a la seva funcionalitat, els robots es poden classificar d'acord amb la tipologia següent:

- **Braços manipuladors** (*manipulators arms*) o **braços robòtics** (*robotic arms*). És el tipus més estès en la indústria (vegeu el subapartat 1.3) i també és molt present als laboratoris de recerca. Són mecanismes poliarticulats, accionats per motors, amb un element terminal en un extrem que pot ser un prensor, mà o pinça per agafar i desplaçar objectes o una eina (de tall, soldadura, pintura, etc.) o fer algun tipus de manipulació. Són habitualment fixos (la base està ancorada al sòl o suspesa d'una estructura), però

Behavioral robotics

També hi ha hagut propostes, com la de la robòtica de comportament (*behavioral robotics*) de Rodney Brooks, que enllaçaven directament percepció-acció en molts comportaments elementals, de l'acció conjunta dels quals emergia un comportament més sofisticat, arguint que el món físic és la seva millor representació. Més endavant aquestes propostes es van decantar per una arquitectura híbrida, en la qual es produïa una certa activitat deliberativa almenys a alt nivell.

Bibliografia

<https://www.iso.org/obp/ui/#iso:std:iso:8373:ed-2:v1:en>

⁽¹⁾ En comparació d'això, un braç robot com el que es descriu a continuació es mou dins el seu espai de treball encara que la seva base estigui fixa.

Sistema cilindre-èmbol

Aquest sistema pot ser una xeringa o els cilindres del motor d'un cotxe. En un actuator pneumàtic l'aire a pressió, en lloc dels dits, s'introdueix per un dels costats del cilindre i fa moure l'èmbol.

PLC

Aquesta sigla correspon a *controlador lògic programable*: és un tipus de miniordinador en què es pot programar una seqüència d'activació, en aquest cas de vàlvules que permeten o interrompen el pas de l'aire als extrems del cilindre movent l'èmbol a un costat o a l'altre.

també poden estar muntats sobre una plataforma mòbil. Un robot com Baxter, dissenyat per a col·laborar amb treballadors humans, consistent en dos braços muntats sobre una columna que conté l'ordinador de control i amb una pantalla que mostra expressions facials, és a la frontera entre els braços manipuladors i els robots humanoides. Més endavant es descriuen els components i principis de funcionament d'un braç robot.

- **Robots mòbils.** La seva característica principal és que són capaços de desplaçar-se d'un lloc a un altre. Solen fer-ho sobre rodes si el sòl és més o menys llis, i sobre potes o erugues si el terreny és abrupte (o sobre rodes muntades en mecanismes que tenen una certa flexibilitat, com els *rovers* d'exploració). També hi ha robots que es desplacen pels medis aquàtic (robots experimentals, d'inspecció i d'exploració) i aeri (vehicles aeris no tripulats, drons). En aquesta categoria s'inclouen des dels AGV (*automated/autonomous guided vehicle*) emprats en logística industrial fins al vehicle autònom, del qual es parlarà extensament en el subapartat 2.2.2.
- **Robots humanoides.** Són robots d'aparença antropomòrfica, encara que aquesta similitud no ha de ser necessàriament molt fidedigna i s'accepten com a tals fins i tot robots que es desplacen sobre rodes o torsos (amb braços i cap) fixos. El més rellevant és que disposi d'extremitats superiors (preferentment dues) i quelcom que es pugui interpretar com un rostre per a facilitar la comunicació. Alguns investigadors posen l'èmfasi en la funcionalitat i desenvolupen robots que es puguin desplaçar sobre cames i agafar coses amb els braços (des del famós Asimo d'Honda fins a l'impressionant Atlas de Boston Dynamics), mentre que altres busquen un aspecte el més humà possible, encara que de moment la motricitat de les extremitats inferiors és molt limitada (per exemple, les *actroid* de la Universitat d'Osaka i Kokoro Ltd, Sophia de Hanson Robotics o les nines i ninots sexuals de RealDoll).
- **Robots experimentals.** En aquest calaix de sastre incloem des de robots biomimètics fins a robots autoreconfigurables, microrobots i nanorobots, robots accionats per cables i altres proves de concepte tant mecàniques com de control.

En un robot cal distingir el maquinari del programari:

- **Hardware:** és la part física del robot, el seu cos, amb el qual interacciona amb l'entorn. Al seu torn, consta de les parts següents:
 - **Estructura.** En el cas d'un braç l'estructura consisteix en una sèrie d'elements rígids connectats entre ells mitjançant **articulacions**, que poden ser rotatives (giratòries, com el colze humà) o lineals (prismàtiques, com el parell cilindre-èmbol). Aquests elements poden estar connectats en sèrie (és a dir, un darrere l'altre com al braç humà) o en paral·lel (com per exemple la plataforma d'un simulador de vol). En el cas d'un robot mòbil l'estructura és la plataforma que conté els motors,

sensors i electrònica, i si és un robot caminant també podem incloure-hi l'estructura de cada pota. De manera més genèrica, l'estructura és el que dona consistència al cos del robot, conté alguns elements com a actuadors i sensors, i permet que el robot es mogui. Es pot establir una analogia anatòmica amb l'esquelet.

- **Actuadors o motors.** Seguint amb l'analogia, corresponen a la musculatura del cos. Són transductors d'energia (generalment elèctrica) en moviment. La majoria són motors elèctrics que activen tant les articulacions d'un braç o d'una cama com les rodes d'un vehicle. També poden actuar en el tancament i obertura d'una mà o pinça, el moviment del cap o de les càmeres, o en els trets facials de robots humanoïdes. Incorporen habitualment algun tipus de reducció mecànica per reduir la velocitat de sortida del motor i augmentar el parell (la força), i un **codificador** (*encoder*), que és un dispositiu que «llegeix» la posició d'un eix (això és, d'una articulació) i la transforma en un senyal digital. Aquests senyals serveixen al controlador del robot per a registrar a tot moment la configuració (la posa) del robot; és el que es coneix com a **propiocepció**. En molts casos és necessari algun tipus de **transmissió mecànica** amb engranatges, corretges dentades o fins i tot sistemes de cables i corrioles, per portar el moviment generat en l'actuador fins a l'articulació que mou (per exemple, en un braç els motors se solen situar a prop de la base per reduir el pes i les forces d'inèrcia del robot).
- **Sensors.** Són dispositius que informen el robot sobre el seu propi estat (per exemple, els codificadors ja esmentats o els mesuradors de càrrega de la bateria) o sobre l'estat de l'entorn. Centrant-nos en aquests últims, poden ser tan simples com un detector de presència o proximitat (per exemple, una cèl·lula fotoelèctrica) que informa sobre la presència d'una peça que pot ser agafada pel robot, o tan complexos com les càmeres d'un sistema de visió, sensors de forces i moments muntats al canell d'un robot (informen sobre les forces exercides al presor del robot) o la pell artificial (matrius de detectors de contacte, de pressió, altres variables com la temperatura, etc.). Un robot també pot incorporar micròfons per a registrar les ordres verbals que li comunica l'usuari.
- **Electrònica.** Comprèn tant l'electrònica de potència (la que regula el subministrament energètic als actuadors) com la de control i l'ordinador (amb els típics components processadors, memòries, disc dur, etc.). L'ordinador s'encarrega del control al més alt nivell, d'interpretar bàsicament els programes emmagatzemats en funció de la informació subministrada pels sensors i, en conseqüència, d'enviar senyals d'activació als diferents actuadors. La presència de l'ordinador és la que confereix la versatilitat al robot, la seva adaptabilitat, sofisticació i facilitat de reprogramació. El robot pot disposar d'una unitat de programació per a introduir els programes, o aquests poden ser escrits en un PC de propòsit general i després transferits a la memòria

de l'ordinador del robot. Als braços manipuladors l'electrònica de potència, l'ordinador i la unitat de programació solen estar separats del braç i integrats en un armari de control. En canvi, en un robot mòbil tant l'electrònica de control com la bateria que subministra energia al robot van a bord.

- **Element terminal.** És el dispositiu que permet al robot fer accions sobre el seu entorn. El més comú és la **pinça** (de dos dits), el **prensor** o la **mà** (aquest terme se sol aplicar quan hi ha més de dos dits) per a agafar i desplaçar objectes per fer-hi un altre tipus de manipulacions. Alguns investigadors intenten emular la versatilitat i destresa de la mà humana, però és un repte que no s'aconseguirà en un futur proper. Tal com s'ha dit abans, altres elements terminals poden ser **eines**: la més utilitzada, per la presència massiva de robots en la indústria automobilitàtica, és la pinça de soldadura per punts, però també són freqüents les eines de tall (convencionals, per làser, plasma o raig d'aigua), les pistoles d'aplicació de pintures i laques, eines d'escatol i poliment, entre d'altres. Els robots d'exploració poden portar dispositius mesuradors i de recollida de mostres.

- **Altres dispositius.** Depenent de les seves característiques, el robot pot incorporar dispositius que facilitin la interacció del robot amb l'usuari, com una pantalla gràfica o un sintetitzador de veu i altaveus per a la comunicació oral, elements estilístics o decoratius, etc.

- **Software:** és la part intangible del robot, la seva **programació**. En la seva forma més simple, consisteix en una seqüència d'especificacions d'accions expressades en el llenguatge de programació del robot, que inclou un repertori limitat d'instruccions de moviment (que s'interpreten, per exemple, com «ves a la posició 12 a la velocitat 0,8»), d'accionament de l'element terminal («obre/tanca pinça») i fins i tot de control del flux del programa (bucles del tipus «mentre hi hagi peces a la cinta transportadora col·loca-les en caixes» o salts condicionals com «si hi ha peça a l'alimentador J, executa les instruccions 14-18; si no, executa les instruccions 56-61»). Tot això en un llenguatge formal propi del fabricant, en el qual paràmetres com «posició 12» o «alimentador J» s'han especificat amb anterioritat, per exemple, guiant manualment el robot a les posicions en qüestió i assignant-los aquests noms. La majoria de robots industrials actuals es programen d'aquesta manera i requereixen a tot estirar formes de percepció simples (per exemple, detectar la presència o absència d'una peça en un lloc especificat). Això és possible gràcies al fet que aquests robots treballen en un entorn controlat, és a dir, on cada cosa és al seu lloc i no hi ha imprevistos. La situació canvia radicalment en el cas de robots de serveis, socials o d'exploració. En aquests casos el robot ha de desenvolupar-se en un món canviant, ple de contingències inesperades, d'informació incompleta o no precisa. És evident que en aquestes circumstàncies el robot requereix un sistema sensorial que li proporcioni una descripció actualitzada

de l'entorn. Si el robot és una aspiradora autònoma que circula per casa, no necessitarà un sistema de sensors molt sofisticat, sino que n'hi haurà prou que detecti obstacles i localitzi el carregador de la bateria. El programari de control podrà equiparar-se al d'un insecte en el millor dels casos. En canvi, si ha d'interactuar amb el propietari, distingir-lo d'altres persones, conèixer les seves preferències i mantenir una conversa estimulante, necessitarà un sistema robust de reconeixement facial (robust en el sentit que sigui capaç de reconèixer-lo en diferents condicions d'il·luminació, punts de vista o opcions de pentinat) i diverses tècniques d'IA de les quals es parlarà en el subapartat 1.4. En qualsevol cas, el fet que el comportament estigui condicionat per programari, sigui codificat per un programador o après per observació és un dels trets distintius del robot. Per tant, un autòmat com els que van ser tan populars entre els segles XVIII i XIX no és un robot malgrat l'aparença humanoide, la sofisticació dels moviments i la naturalesa mecànica, ja que per a variar la tasca que fa serà necessari introduir modificacions físiques en el mecanisme, com canviar el sistema de lleves que regeix els moviments de les mans de la clavecinista dels germans Droz. En canvi, un parent proper són les màquines de control numèric o les impressores en 3D, que creen peces físiques a partir de les especificacions d'un programa.

1.3. Aplicacions

Els àmbits d'aplicació dels robots són diversos i variats. El sector dominant és l'industrial per motius històrics i econòmics. Però, a poc a poc, van penetrant en altres sectors i fent tasques que semblaven reservades als humans. Repassem-les breument:

- **Robots industrials.** És el sector pioner: el 1954 la invenció del primer «aparell programat per a transferir articles» de George Devol, que el 1956 va conèixer l'empresari Joseph Engelberger, va marcar-ne l'inici quan a l'any següent tots dos van desenvolupar el primer prototip, l'Unimate 001 (Robotic Industries Association, 2019). El 1961 Engelberger va fundar Unimation Inc., la primera empresa dedicada a fabricar robots industrials. Des de l'inici el principal client de les companyies de fabricació de robots industrials va ser la indústria automobilística, que avui dia encara lidera el sector, seguida molt de prop per la indústria d'aparells i components electrònics. Això es deu al fet que al seu moment la indústria automobilística tenia prou envergadura (i beneficis) per a poder permetre's apostar per una tecnologia nova els resultats de la qual encara es desconeixien. La producció en massa dels seus productes i el tipus de tasques encomanades als robots (tasques perilloses en ambients poc saludables, penoses i monòtones)² van propiciar la substitució progressiva de treballadors humans a les línies de soldadura i pintura. Segons dades de la International Federation of Robotics (IFR), el 2018 la venda de robots va ser de 384.000 unitats (World Robotics, 2019), la qual cosa situa el nombre total de robots industrials

Robot industrial

La mateixa norma dona definicions per a alguns dels termes. Així, *reprogramable* significa «dissenyat de manera que els moviments o funcions auxiliars programats poden ser modificats sense alteració física»; *multipropòsit* és «capaç de ser adaptat a una aplicació diferent amb alteració física» (per exemple, un canvi d'element terminal, que també pot ser automàtic), i *manipulador* és una «màquina on el mecanisme consisteix usualment en una sèrie de segments, articulats o que llisquen un respecte a l'altre amb el propòsit d'agafar i/o moure objectes (peces o eines) generalment en diversos graus de llibertat».

⁽²⁾ És el que es coneix com la triple D en anglès: *dangerous, dirty and dull*.

instal·lats al món en prop de 2,5 milions d'unitats (World Robotics, 2018). La Norma ISO 8373:2012 proporciona una definició precisa de robot industrial:

«Manipulador controlat automàticament, reprogramable, multipropòsit, programable en tres o més eixos, que pot ser fix en un lloc o mòbil per a ús en aplicacions d'automatització industrial.»

Les tasques en què s'empren robots industrials inclouen soldadura per punts i soldadura contínua, pintura, tall, acabat (escatat, poliment, encerat, etc.), assemblatge o desassemblatge, paletitzatge, alimentació de màquines (això és, transferència de peces a màquines i de màquines), transport i logística. Les operacions de desassemblatge en particular són de gran interès per a la incipient indústria del reciclatge i recuperació de components o metalls valuosos. Un altre repte interessant, amb grans dificultats però també grans beneficis potencials, és la manipulació d'objectes deformables com menjar i roba.

- **Robots en mineria.** Així com la tecnologia robòtica està més que provada i és operativa en la indústria, els robots encara estan en fase experimental en el sector primari. No obstant això, els avantatges d'una futura robotització són evidents, sobretot en el sector de la mineria, on pot salvar vides. S'estima que uns 12.000 miners moren en accident cada any al món, sobretot en mines de carbó (Chakravorty, 2019). Hi ha diversos projectes de recerca sobre l'exploració de mines inundades, combinació de xarxes denses de sensors i robots per a explorar i monitorar mines, o col·laboració en equips formats per humans i robots en mineria a la Colorado School of Mines i en el Council for Scientific and Industrial Research de Sud-àfrica. L'ús de robots, a més de fer possible l'explotació de mines en presència de gasos tòxics o amb risc d'esfondrament, permetrà reconsiderar l'explotació de mines per a cercar metalls rars la demanda dels quals, abans inexistent, creix als últims anys. També la mineria de superfície pot beneficiar-se de robots de trepatge (menys invasius que els explosius) i el transport de mineral en camions autònoms (Gendron, 2019).
- **Robots en agricultura, ramaderia i pesca.** En agricultura les aplicacions (algunes ja en fase d'explotació) inclouen l'ús de drons i imatges aèries per a avaluar l'estat dels cultius, l'aplicació localitzada d'herbicides i pesticides amb drons o amb robots terrestres, l'eliminació mecànica de mala herba amb robots (prescindint d'herbicides), la navegació autònoma de tractors, la recol·lecció robotitzada de fruita, la collita en plantacions d'hivernacle o la manipulació de plantes d'hivernacle (Postscapes, 2019). En ramaderia podem esmentar un robot teleguiat per a conduir bestiar en recintes barrats (Cargill, 2018), els sistemes robotitzats de munyció (Michigan State University, 2019) i, de manera més experimental, sistemes multirobot per a pasturar ramats d'ovelles (Pierson, 2015). I en pesca es pot fer referència, a manera de curiositat, als anomenats esquers robòtics, que tenen l'aspecte de peixets i n'imiten els moviments aleatoris, per a actuar com a esquers

Projectes de recerca

Projecte Europeu UNEXMIN, <https://www.unexmin.eu/>
Projecte ARIDuA, Autonomous Robots and the Internet of Things in Underground Mining, <https://tu-freiberg.de/en/aridua/summary>

Més informació sobre els esquers robòtics:

<https://roboticlure.com/>

de peixos majors, encara que no són realment robots perquè no perceben l'entorn ni actuen en conseqüència.

- **Robots d'exploració.** Són robots mòbils equipats per a explorar i agafar mostres o fer alguna altra tasca en entorns hostils per a l'ésser humà, com les profunditats submarines, els volcans, les zones afectades per una catàstrofe o l'espai exterior (inclosos altres astres). Porten generalment una o diverses càmeres, que permeten a l'operador humà visualitzar l'entorn immediat del robot. També poden portar instrumental especialitzat per a la tasca específica que han de dur a terme. Per exemple, els robots de cerca i rescat, la missió dels quals és localitzar persones amb vida entre els enderrocats o en condicions de mala visibilitat, estan equipats amb sensors infrarojos passius que detecten fonts de calor. També han d'estar equipats per a resistir durant un temps raonable en les condicions ambientals adverses en què desenvolupen la seva activitat, com pressions i/o temperatures extremes, radiació, microimpactes, etc. Els robots d'exploració submarina (Smithsonian Ocean Team, 2009), volcànica (Muscato *et al.* 2012) o de zones afectades per radiacions (González, 2015) solen ser unitats teleguiades, ja que la distància és relativament curta i l'operador del robot pot donar resposta immediata a les situacions captades per les càmeres de bord. Les dificultats tècniques rauen sobretot en la protecció de les parts electròniques més sensibles a les condicions extremes en què operen. No obstant això, les autèntiques estrelles de l'exploració són els robots espacials i en particular els *rovers* d'exploració planetària enviats a Mart. El 1970 i el 1973 els soviètics ja havien enviat *rovers* d'exploració a la Lluna amb el programa Lunokhod. Aquestes unitats operaven remotament, la qual cosa no era factible per als *rovers* marcians a causa del gran retard del senyal. El primer va ser el Mars Pathfinder (1997), al qual van seguir l'Spirit i l'Opportunity (van arribar a Mart el 2004) i el Curiosity (2012) (Mars Exploration Program, 2019). Aquests robots tenen certa autonomia per a determinar la direcció a seguir o evitar un obstacle, ja que si depenguessin enterament de les instruccions enviades des de la Terra avançarien molt a poc a poc. Tots porten càmeres per a enviar vídeos i fotografies a la Terra i un equip científic per a agafar i analitzar mostres (els resultats de les quals també són enviats a la Terra). Els robots desactivadors de mines també es poden considerar robots d'exploració, però s'han inclòs en l'apartat de robots militars pel context bèl·lic o postbèl·lic en què se situen (vegeu més endavant «Robots militars» i el subapartat 2.2.3).
- **Robots de construcció i manteniment.** Els robots de construcció amb un futur immediat més prometedori són les impressores en 3D d'edificis, robots que col·loquen maons i robots de demolició (Robotics Online, 2018). Pel que fa al manteniment, ja hi ha robots de neteja de façanes (generalment grans superfícies de vidre) o d'inspecció i neteja de canonades, entre d'altres.

- **Robots de transport de persones o mercaderies.** Aquesta categoria inclou els vehicles autònoms terrestres i els vehicles aeris no tripulats, dels quals es parlarà extensament en el subapartat 2.2.2. Molts tenen l'aparença de vehicles convencionals (cotxes, camions, autocars), però van equipats amb un complex sistema sensorial.
- **Robots militars.** En el subapartat 2.2.3 es descriuen breument la tipologia d'aquests robots d'acord amb el tipus de missions encomanades i, de manera més extensa, els vehicles aeris no tripulats (UAV). Des del punt de vista constructiu, a més dels UAV, hi ha també sistemes amb armes que es reorienten autònomament sobre una base fixa a gran velocitat per repel·lir atacs, i armes o manipuladors (per a desactivar explosius) muntats sobre *rovers*, generalment teleguiats.
- **Robots de serveis.** La Norma ISO 8373:2012 defineix el robot de servei com un «robot que fa tasques útils per als humans o l'equipament, excloent aplicacions d'automatització industrial». Segons aquesta definició, és com un calaix de sastre en què té cabuda tot tipus de robots excepte els específicament industrials. No obstant això, en la pràctica robòtica se solen considerar els que operen en un exterior no urbà com a «robots de camp» (*field robots*), mentre que als robots de serveis se'ls atribueix un component social major, d'interacció amb els humans. La frontera respecte a les altres categories és difusa; per exemple, els robots de neteja en edificis o als carrers, que no deixen de ser robots de manteniment, es poden considerar perfectament robots de serveis. Aquí també entren els robots emprats en hostaleria (cambrers, cuiners i servei d'habitacions) i en centres educatius, hospitals, etc., que exerceixen tasques similars. També es consideren robots de serveis els informatius i guia per a visitants i turistes en entorns urbans, museus, congressos, fires, parcs, etc. Depenent de les tasques encomanades, aquests robots han d'estar molt bé equipats des del punt de vista sensorial i perceptiu, i també molt desenvolupats en aspectes d'interacció amb els éssers humans (reconeixement facial i d'emocions, interpretació del llenguatge natural o almenys de certes frases i gestos, síntesi de veu, *displays* gràfics informatius, entre d'altres). Per a una revisió de reptes i assoliments recents per a aquest tipus de robots, vegeu Torras (2016). Precisament, els robots que posen èmfasi en la interacció amb els éssers humans constitueixen la categoria de **robots socials**, en els quals també entren els de l'apartat següent. Vegeu Torras (2018b) per a una revisió de les implicacions ètiques associades a aquest tipus de robots i en les xarxes socials, i el paper especulatiu (i educatiu) de la ciència-ficció en els avantatges i inconvenients de possibles escenaris relacionats amb aquestes tecnologies.
- **Robots domèstics i personals.** Tenen característiques molt similars a les exposades en l'apartat anterior, però en aquesta categoria es posa èmfasi en l'adaptació a l'usuari (o grup reduït d'usuaris, com una família, per exemple). Aquests robots han de tenir una gran capacitat d'aprenentatge i venen

de fàbrica amb una sèrie d'habilitats manuals i socials. Entren en aquesta categoria tots els robots que es mouen en l'àmbit domèstic, la qual cosa inclou des de senzills robots de neteja fins a robots més sofisticats que actuen com a majordom o com a secretari personal, robots d'entreteniment i fins i tot companys sexuals o sentimentals. Els riscos que aquests robots comporten per a la privadesa de l'individu s'analitzen en el subapartat 2.2.5, i els riscos per al seu equilibri emocional en el subapartat 2.2.6.

- **Robots educatius i d'entreteniment, «mainaderes» robòtiques.** Són pensats per a educar de manera pràctica en robòtica o en altres matèries, o per a actuar com a joguines sofisticades, i poden incloure la possibilitat de ser assemblats o reconfigurats pels usuaris. Es poden utilitzar tant en l'àmbit domèstic com en centres escolars, on actuen com a element de suport al professorat. Tenen l'avantatge de poder dedicar una atenció exclusiva i ininterrompuda a cada infant, tenen una paciència i una constància il·limitada (qualitats molt apreciades en aquest context) i al mateix temps una fermesa immune al xantatge emocional. Són molt atractius per als petits, la qual cosa facilita que s'involucrin més en l'aprenentatge. També contribueixen al fet que els infants puguin confiar que no seran criticats pel robot si s'equivoquen. Poden constituir una eina inclusiva per als infants amb dificultats de socialització si s'usen correctament, però també poden produir justament l'efecte contrari. Entre els efectes negatius potencials hi ha una disminució de la tolerància a les discrepàncies dels altres, una dependència excessiva, l'atribució de desitjos, emocions, objectius propis i afectes a la màquina (i la consegüent decepció en adonar-se que no hi són), i l'exposició de l'infant a la possible activitat d'un *hacker*. Alguns robots d'entreteniment també poden fomentar la passivitat i la falta d'imaginació. El col·lectiu infantil és molt vulnerable i mal·leable, per la qual cosa el disseny i la programació d'aquests robots ha de posar especial èmfasi en la seguretat i en la pedagogia.
- **Robots mèdics, de rehabilitació i assistencials.** Els robots emprats en medicina fan operacions de cirurgia, assisteixen en labors d'infermeria amb tasques rutinàries com distribució de medicaments o bé prestant assistència física al moviment de pacients, i poden ser emprats també en teràpies de rehabilitació. A més de prestar serveis en hospitals, poden complir el paper d'assistència domiciliària de pacients impedits o d'edat avançada, i en aquest cas constitueixen una categoria especial dels robots personals i domèstics (de nou, vegeu els subapartats 2.2.5 i 2.2.6 per a algunes consideracions ètiques de seguretat i dependència emocional). La majoria de robots cirurgians són en realitat telerobots, que reproduïxen, a una escala molt menor i evitant possibles tremolors, els moviments fets pel cirurgià humà des de la seva interfície. El cirurgià pot seguir el progrés de l'operació gràcies a les imatges mostrades per un monitor i capturades per microcàmeres introduïdes en el pacient o a altres sistemes de captura d'imatges. Això permet augmentar el camp visual, facilita una experiència immersiva del cirurgià en l'entorn anatòmic de l'operació miniaturitzant-ne els mo-

viments i aconsegueix, per tant, una major precisió. A més, el cirurgià pot predefinir zones prohibides, teixits com per exemple nervis o alguns vasos sanguinis que no s'han de tocar, i el sistema de control impedeix l'accés a aquestes zones en temps real. L'ús de robots sol estar associat a cirurgia mínimament invasiva, i l'exemple més conegut és el robot Da Vinci (Da Vinci, 2019), que ja s'usa en nombrosos hospitals. Els robots que treballen de manera totalment autònoma, com el robot dentista que va fer dos implants el 2017 (Yan, 2017), encara estan en fase experimental, com també ho estan els microrobots i nanorobots, que s'introduïrien en el cos del pacient per actuar *in situ* (destruint un coàgul o un tumor, per exemple). Els robots auxiliars d'infermeria són versions hospitalàries de robots de serveis més comuns, excepte els que han d'entrar en contacte físic amb el pacient per a moure'l. Àdhuc actuant sota supervisió del professional sanitari, aquests robots han de complir requisits de seguretat estrictes, presentar una estabilitat elevada, un disseny ergonòmic per a una interacció física més adequada i capacitats sensorials precises. La seva introducció als hospitals millorarà en gran manera les condicions laborals del personal d'infermeria, exposat al risc de lesions en desplaçar pacients voluminosos. Quant als robots d'assistència terapèutica i rehabilitació, es tracta de sistemes capaços d'ajustar les teràpies d'acord amb les característiques i el progrés del pacient. Els exercicis en les variants de reforç de l'aparell locomotor, coordinació, destresa, memòria, etc. són proposats i supervisats pel terapeuta humà, però és el robot que pot augmentar o disminuir el grau de dificultat en funció de la resposta del pacient. A més de descarregar el terapeuta del treball més rutinari, el pacient es pot sentir més confiat perquè no és sotmès a eventuais crítiques pel sistema.

- **Exoesquelets i pròtesis robòtiques.** De les pròtesis robotitzades i de les seves implicacions ètiques es parlarà extensament en el subapartat 2.2.4. Quant als exoesquelets, es tracta d'estructures que s'ajusten al cos de l'usuari, són dotades de sensors que poden captar certa activitat muscular o neuronal del pacient i donar una resposta en forma de moviment mitjançant els seus actuadors. Poden amplificar la força física d'un usuari sa (de fet, s'està investigant el seu possible ús en aplicacions militars), però sobretot poden proporcionar mobilitat a un usuari impedit.

1.4. Tècniques d'intel·ligència artificial en robòtica

La majoria de robots existents avui dia actuen d'acord amb una seqüència d'instruccions codificada per un programador humà. Tal com s'ha dit, això és suficient per a robots que exerceixen l'activitat en un entorn controlat, com els robots industrials, o que fan tasques simples, com un robot aspirador. Quan el robot s'enfronta a un món complex, dinàmic, observable parcialment i ple d'incerteses, ha de ser capaç de perseguir els objectius marcats malgrat obstacles i contingències, i ha d'afrontar les ambigüitats i matisos de les relacions

humanes, i no hi ha dubte que solament pot fer-ho amb èxit si està dotat de les eines cognitives de l'IA. A continuació explicarem succintament en què consisteixen aquestes eines; per a més detalls, vegeu Jiménez (2018):

- **Percepció.** És l'adquisició, processament i interpretació, en termes intel·ligibles per al sistema de representació del coneixement i control del robot, dels senyals visuals, de forces i moments, acústics, tàctils, etc. que proporcionen els sensors. És la premissa per als processos d'aprenentatge i control d'un robot, la qual cosa li permet estar situat al món i donar una resposta coherent amb el que interpreta que passa al voltant. El canal sensorial més informatiu, encara que també més complex, és la visió per computador. Les imatges adquirides per la càmera són segmentades en diferents regions significatives, i sobre aquesta partició tenen lloc els diferents processos de detecció (per exemple, d'ulls i boca en una cara), d'identificació (per exemple, associant noms a diferents cares) o de reconeixement (interpretant una escena com a «dona muntada amb bicicleta amb un gos que corre al costat»). La integració de la informació sensorial de diferents canals (visió i tacte, per exemple) permet resoldre ambigüitats dels canals per separat i es denomina **fusió sensorial**.
- **Aprenentatge.** Consisteix a adquirir o actualitzar un model del món de primera mà, això és, processant la informació que subministra el sistema sensorial. Les tècniques de base són mètodes matemàtics de correlació, probabilístics, estadístics o de reconeixement de patrons. Les tècniques d'aprenentatge es poden aplicar a la interpretació d'imatges i a la interpretació del llenguatge natural, assignant estadísticament un significat a cada element de la imatge o del text. Un mètode que té gran èxit són les xarxes neuronals artificials ja esmentades (en el subapartat 1.1) en una versió actualitzada que inclou moltes capes de neurones (les anomenades *deep neural networks* que fan possible el famós aprenentatge profund), encara que tenen el gran inconvenient de la falta de traçabilitat (és un procés del tipus «caixa negra», en què no se sap què passa a l'interior i per què aprèn el que aprèn). L'aprenentatge pot ser **supervisat** (per un humà, per exemple) mitjançant tècniques de classificació (el supervisor indica la classe a què pertany cada mostra d'aprenentatge o entrada) o de regressió (ajustament automàtic d'una funció matemàtica als parells entrada-sortida, donada pel supervisor), o **no supervisat**, amb aplicació de tècniques de clusterització (agrupació per similitud) o d'aprenentatge de regles associatives. Els dos grans paradigmes d'aprenentatge en robòtica són l'**aprenentatge per reforç** (*reinforcement learning*), en què el robot aprèn quines accions són més convenientes en cada estat de l'entorn mitjançant observació i una funció de recompensa (o penalització), i l'**aprenentatge per demostració**, en què un «professor» humà ensenya al robot la millor manera de fer una acció o una tasca. L'aprenentatge per reforç té l'avantatge que permet al robot aprendre autònomament, encara que pot allargar-se molt, ja que ha d'anar provant les diferents accions en entorns canviants. Això no passa amb l'aprenentatge per demostració, que té, no obstant això, l'inconvenient de

requerir repetides demostracions del professor (per a adquirir certa generalitat davant lleugeres variacions en l'entorn i aprendre solament el que és realment significatiu), la qual cosa pot ser enutjós o tediós per a ell. Ambdues tècniques es poden combinar: el professor, amb poques demostracions, pot deixar que el robot aprengui per reforç pel seu compte i millori així les accions apreses recentment.

- **Planificació.** És determinar la seqüència de moviments o accions necessària per a aconseguir un objectiu. Cal distingir la planificació de moviments de la de tasques. La **planificació de moviments** és purament geomètrica (encara que també pot incloure velocitats i acceleracions) i consisteix a determinar la seqüència de poses (posició i orientació) d'un robot per a arribar d'un punt d'inici a una meta evitant els obstacles (fixos o mòbils) que pugui haver-hi al seu entorn a partir d'una representació geomètrica explícita o implícita d'aquest entorn. La **planificació de tasques** opera en un àmbit simbòlic amb un repertori d'accions bàsiques que el robot pot executar. Aquestes accions solament es poden executar si es donen certes precondicions i, una vegada executades, produeixen certs efectes en l'entorn, modificant-lo de manera que ara es donen les precondicions d'altres accions diferents. Això permet planificar, això és, encadenar les accions bàsiques de manera que es vagi modificant l'entorn fins a aconseguir l'estat desitjat. A l'inici els mètodes de planificació eren deterministes, això és, partien de la base que tota la informació rellevant estava disponible i que els efectes de les accions sobre l'entorn eren fixos i determinats. Avui dia hi ha planificadors capaços de proporcionar un bon pla en presència de models incomplets del món i incertesa en els efectes de les accions.
- **Representació del coneixement i raonament.** Juntament amb representacions elementals com a conjunts de dades, descripcions geomètriques, grafs (i les corresponents tècniques de cerca en grafs), etc., el coneixement pot ser representat amb la **lògica**, amb la qual les afirmacions tenen un valor binari (veritable o fals) i que permet raonar per inferència inductiva, deductiva o abductiva. Una manera molt efectiva d'utilitzar la lògica és mitjançant la programació de satisfacció de restriccions (*constraint logic programming*). D'altra banda, la lògica de descripcions (*description logics*) parteix d'una estructura terminològica jerarquitzada i utilitza la consistència, subsumpció (en la seva accepció més general, no jurídica) o disjunció com a mecanismes de raonament (entre d'altres). Està en la base de l'OWL (*ontology web language*) per a representar el coneixement a internet (per a ser usada per aplicacions més que per humans) i també de RoboEarth, un repositori en xarxa per a robots (vegeu el subapartat 2.2.5). Una altra forma de representació és la **probabilística**, més d'acord amb les incerteses del món real, on les afirmacions són certes o falses amb certa probabilitat i el raonament es fonamenta en el teorema de Bayes per a determinar la probabilitat d'una hipòtesi atesa l'evidència observada (s'utilitzen formalismes com les xarxes bayesianes). Finalment, en la **lògica difusa** (*fuzzy logic*) les afirmacions tenen un determinat grau de certesa: diferents afir-

Planificació de tasques

Aquesta planificació és amb representacions simbòliques de les accions, per exemple, en forma de regles com «si es donen aquestes *precondicions*, llavors l'acció pot executar-se i produeix aquests *efectes*».

macions poden ser certes simultàniament, però en cada situació unes ho són més que unes altres.

Un repte que encara és lluny de poder-se aconseguir és la recreació artificial de consciència i emocions. La primera té a veure amb la consciència d'un mateix (*self-awareness*) que mostren molts animals, sobretot mamífers superiors, amb les experiències subjectives (*qualia*), que alguns filòsofs i científics consideren el principal obstacle per a aconseguir una consciència genuïna, amb la voluntat i els objectius propis, etc. Cal preguntar-se si l'objectiu d'aconseguir una consciència artificial, al marge de proporcionar nova llum sobre què és la consciència humana, és desitjable per si mateix, i si és convenient o no dotar les màquines de consciència. En una línia similar hi ha les emocions. Per descomptat, pot ser molt útil que el robot pugui arribar a identificar, mitjançant l'aprenentatge, emocions i l'estat d'ànim de les persones amb les quals ha d'interaccionar, de manera que moduli aquesta interacció de la manera més convenient (per exemple, no fent-se pesat). Fins i tot el fet que el robot pugui simular certes emocions (per exemple, mostrar-se cansat si està baix de bateries) pot facilitar la interacció. Però les possibilitats i la utilitat que algun dia una màquina senti emocions realment són més que qüestionables.

2. Consideracions ètiques

Els avenços en IA i robòtica i els productes que es deriven d'aquest desenvolupament tenen un impacte potencial de gran importància en la societat humana. Altres tecnologies, com per exemple internet o la telefonia mòbil, ja han donat proves del seu poder transformador en la manera que tenim de comunicar-nos, comprar, viatjar, consumir productes culturals o relacionar-nos. Fins i tot si es compleixen únicament les expectatives dels més escèptics (o els menys endevins i malastrucs), no hi ha dubte que els robots afectaran de diverses maneres la nostra manera de viure conforme vagin guanyant prestacions, destresa, autonomia i habilitats cognitives. La mateixa comunitat científica ja va donar els primers passos en considerar les implicacions ètiques d'aquesta tecnologia. El començament oficial de la **roboètica** o ètica del desenvolupament dels robots per al progrés humà i social, terme encunyat pel científic i expert en robòtica Gianmarco Veruggio el 2002, se situa a l'any 2004 amb el Primer Congrés de Roboètica a San Remo (Itàlia). Aquest mateix any es va crear el Comitè Tècnic en Roboètica de l'IEEE-RAS i va tenir lloc la Declaració Mundial del Robot a Fukuoka (Japó) sobre les expectatives dels robots de propera generació (Fukuoka, 2004).

En aquesta part repassarem alguns dels àmbits de l'activitat humana en els quals l'IA i la robòtica poden tenir una incidència més acusada, i les qüestions d'índole ètica i moral que cal plantejar-se sobre els possibles riscos que es deriven de la seva implantació. A la curta, aquests riscos no difereixen substancialment dels derivats del mal funcionament de qualsevol altra màquina, inclosos els errors mecànics i humans. A la llarga, i això és mera especulació, els riscos són els associats a l'existència d'una nova espècie intel·ligent a la Terra, amb la salvaguarda d'una eventual fusió entre la humanitat biològica i les seves criatures cibernètiques. I, entremig, els riscos provenen de la presència ubíqua de màquines que permearan cada vegada més sectors i faran cada vegada més coses; moltes d'aquestes coses les faran millor que els seus creadors humans, i per tant aniran guanyant més atribucions i responsabilitats. A continuació utilitzarem freqüentment ètica i moral com a termes intercanviables, encara que el lector ha de ser conscient de les diferències entre totes dues: mentre que la moral es refereix al conjunt de costums imposats per l'entorn, als usos d'una societat determinada en el temps i el lloc, l'ètica és fruit d'una reflexió individual, racional, i influeix en la conducta personal de manera voluntària i conscient. La moral té un caràcter més pràctic i dona resposta a situacions concretes, viscudes en el dia a dia, mentre que l'ètica intenta donar resposta des d'un punt de vista teòric a per què hauríem d'actuar (o no) de la manera que proposa la moral (Román, 2016).

Primer Congrés de Roboètica

http://www.roboethics.org/sanremo2004/ROBOETHICS_Program.html

IEEE

És l'associació internacional d'enginyers elèctrics i electrònics més important del món, de la qual forma part RAS (Robotics and Automation Society). <http://www.ieee-ras.org/robot-ethics>

2.1. És possible construir màquines ètiques?

Aquesta pregunta és prou inconcreta per a admetre diverses interpretacions i, per tant, diverses respostes. En primer lloc, arran del que hem vist fins ara, sabem que els robots són màquines rellevants èticament en la mesura que poden ocasionar amb les seves accions diferents efectes que afecten tant particulars com sectors i el conjunt de la societat, en forma de danys físics, econòmics, psicològics o morals. En segon lloc, els robots no estan en condicions de respondre per les seves accions donat l'estat actual de l'IA. El robot que va causar l'accident mortal de la planta d'Akashi i va ser retirat no *va patir* realment un càstig i podria haver continuat funcionant igual que les altres unitats; la decisió de retirar-lo va respondre a la compulsió humana del rescabament, de la mateixa manera que un infant «castiga» donant un cop al moble contra el qual acaba de topar. Per tant, a dia d'avui el que tenim són agents (màquines) ètics *implícitament*, en terminologia de James Moor (2006), o màquines dotades de **moralitat operacional**, en paraules de Wallach i Allen (2009): màquines la

«significació moral de les quals està enterament en mans de dissenyadors i usuaris» (Allen i Wallach, 2014).

Per exemple, a dia d'avui els vehicles autònoms no van més enllà d'aquesta modalitat segons Loh i Loh (2017). Malgrat constituir (junt amb el conductor) el que aquests autors denominen un sistema híbrid pel que fa a prendre decisions durant la conducció, la responsabilitat en situacions de dilema moral recau en el conductor. En contraposició a això, hi ha les màquines que són *agents ètics explícits* les accions dels quals són fruit d'un raonament explícit i deliberat sobre els efectes ètics d'aquestes accions. Per a Wallach i Allen, la màxima expressió de l'agent ètic explícit se situa en el que denominen **agència moral plena** (*full moral agency*), característica ara com ara exclusiva de l'ésser humà. No és descartable que algun dia les màquines puguin adquirir quelcom semblant a la consciència, a la voluntat i als mateixos objectius (encara que cal preguntar-se si això és necessari o desitjable); aquest desenvolupament vindrà guiat possiblement per l'IA general o forta.

Mentrestant, Wallach i Allen postulen la possibilitat que els robots puguin tenir el que ells anomenen **moralitat funcional**, amb la qual, fruit d'una activitat deliberativa (explícita), «les mateixes màquines tindran la capacitat per a avaluar i respondre qüestions morals». És a dir, al moment de decidir la següent acció a executar (o de planificar una seqüència d'accions), una màquina funcional moralment ha de ser capaç d'avaluar el valor ètic dels efectes d'aquesta acció i determinar si aquests efectes (i, per tant, l'acció mateixa) són acceptables o no a la llum d'aquesta avaluació. La moralitat funcional se situa entre els extrems de la moralitat operacional i l'agència moral plena. Per als autors (i, de fet, hi ha un ampli consens entorn d'aquesta qüestió), aquesta capacitat no és solament desitjable sinó necessària conforme els robots gua-

nyin competències i atribucions. La pregunta següent és: com s'implementa aquesta moralitat funcional? O, en termes de Wallach i Allen, com es construeix un *agent moral artificial* (AMA)?

Fa segles que els filòsofs debaten sobre ètica. Han desenvolupat abstraccions, teories i sistemes amb pretensions d'universalitat, que transcendeixen fets i situacions concretes però que per aquest mateix motiu s'haurien de poder aplicar en qualsevol circumstància i context particular. D'altra banda, els enginyers i investigadors en IA i robòtica estan interessats precisament en l'aspecte pràctic del disseny i implementació d'un sistema que funcioni en circumstàncies i contextos particulars. Poden aprofitar els assoliments de la filosofia i evitar haver de reinventar la roda? Per a respondre aquesta pregunta, hem de repassar les principals teories de l'ètica i examinar les seves possibilitats d'implementació en una màquina:

1) La **deontologia** es refereix a l'estudi dels fonaments del deure i les normes morals. El postulat bàsic de la deontologia és que les accions tenen un valor per si mateixes i que hi ha deures que s'han de complir independentment de les seves conseqüències. En la pràctica això es tradueix en l'establiment de conjunts ordenats de regles que regeixen el comportament moral d'un col·lectiu, per exemple, dels qui exerceixen una determinada professió (deontologia professional). El representant més destacat d'aquesta teoria ètica és probablement Immanuel Kant (1724-1804) i el seu imperatiu categòric. L'ambigüitat inherent al llenguatge humà fa possible un cert marge d'interpretació i el debat sobre si aquesta ha de cenyir-se al text o a l'esperit de les regles. La deontologia evoca immediatament les famoses tres lleis de la robòtica de l'escriptor de ciència-ficció Isaac Asimov (1920-1992), de les quals parlarem més endavant.

2) El **conseqüencialisme** adopta la perspectiva oposada en la mesura que no valora les accions per si mateixes ni en la intenció que les anima sinó per les seves conseqüències. Per tant, una acció serà bona si ho són les seves conseqüències. Una de les variants més destacades del conseqüencialisme, l'**utilitarisme**, desenvolupat formalment per Jeremy Bentham (1748-1832) –encara que el terme va ser encunyat pel seu deixeble John Stuart Mill (1806-1873)–, se sol resumir en el principi d'utilitat d'aconseguir «el màxim benefici per al màxim nombre de persones». Aquest *utilitarisme de l'acte* se sol contraposar a l'*utilitarisme de les normes*, i aquest últim defineix la millor acció com la que segueix una norma que aporta la màxima utilitat. L'utilitarisme de les normes admet en certa manera una conciliació del conseqüencialisme amb la deontologia. Altres variants del conseqüencialisme inclouen l'egoisme ètic, l'altruisme ètic o el conseqüencialisme (o utilitarisme) negatiu (que propugna concentrar-se a reduir el sofriment més que a augmentar el plaer).

3) L'**ètica de les virtuts** està relacionada amb el caràcter d'una persona: una acció és correcta èticament si és la que faria una persona virtuosa en aquesta situació. Té l'origen en Plató i Aristòtil, i els conceptes que encara avui dia manegen les versions actualitzades d'aquest corrent provenen de la filosofia grega

i inclouen l'*areté* (excel·lència o virtut), la *phronesis* (saviesa pràctica o moral, que permet a l'individu saber fer allò correcte en cada situació) i l'*eudaimonia* (felicitat, un estat que s'aconsegueix precisament per la pràctica de les virtuts). Vegeu Wikipedia, «Virtue ethics» (2019), per a una llista d'autors moderns que subscriuen l'ètica de les virtuts, i de les virtuts aristotèliques i els vicis que resulten del seu excés o del seu defecte.

¿És possible implementar alguna d'aquestes teories en un robot de manera que sigui capaç de passar les seves accions pel tamís de l'ètica abans d'executar-les? La ciència-ficció ja va proposar una manera de fer-ho als anys quaranta del segle passat mitjançant les famoses tres lleis de la robòtica, que apareixen per primera vegada en el relat «Cercle viciós» («Runaround», en el número de març de 1942 de la revista *Astounding Science Fiction*), de l'escriptor Isaac Asimov, i posteriorment en la col·lecció de relats *Jo, robot* (1950), entre d'altres:

- «1. Un robot no ha de fer mal a un ésser humà ni permetre, per inacció, que un ésser humà pateixi un dany.
2. Un robot ha de complir les ordres donades pels éssers humans a excepció de les que entrin 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 segona llei».

Posteriorment, s'hi afegiria una «lleï zero», a la qual van supeditades les anteriors:

- «Un robot no farà mal a la humanitat ni permetrà, per inacció, que la humanitat pateixi un dany.»

En aquella època eren molt populars les novel·les de gènere en què els robots es rebel·laven invariablement i constituïen una amenaça a l'altura de l'heroïcitat del protagonista. Asimov considerava absurd que el desenvolupament hipotètic de robots intel·ligents no comportés una salvaguarda per als seus creadors humans. Les tres (o quatre) lleis formaven part de la mateixa estructura dels cervells *positrònics* dels robots, de manera que no podien ser *hackejades*. Evidentment, Asimov no proporciona detalls sobre la construcció d'aquest tipus de cervells artificials ni pretenia que aquestes lleis s'implementessin d'alguna manera en màquines reals: no eren més que un recurs literari que li permetia desenvolupar les trames associades a situacions en què són un obstacle en un conflicte. No obstant això, se'ls pot reconèixer el mèrit de constituir un primer codi deontològic per als robots, encara que sigui en la ficció. Diverses associacions professionals, universitats i altres institucions han desenvolupat les seves pròpies llistes de principis ètics de la robòtica i l'IA, un compendi de les quals es pot trobar a Winfield (2018).

Les teories normatives de l'ètica, siguin deontològiques o conseqüencialistes, procedeixen de manera descendent (*top-down*): un principi genèric, sigui l'adhesió a un deure expressat en una regla o la maximització d'un benefici, ha de ser aplicat a una situació concreta. En robòtica i en IA és freqüent programar

instruccions en forma de regles que han de ser instanciades per a poder executar-se (això és, la informació que procedeix dels sensors indica si es donen les condicions a l'entorn per a executar l'acció expressada en la regla) o funcions amb variables que també corresponen a valors determinats pels sensors. No obstant això, aplicar aquests procediments a l'avaluació ètica d'una acció en un entorn poblat d'humans (més enllà d'un entorn restringit i controlat com l'industrial) no és tan immediat. En el cas de les formulacions deontològiques, el primer problema amb què ens trobem és el de la interpretació correcta de la regla en qüestió. Una interpretació literal de la primera llei d'Asimov impulsaria el robot a detenir la mà del cirurgià que està a punt de practicar una incisió. La casuística d'excepcions a una regla genèrica pot ser aclaparadora. També ho és la de possibles realitzacions o concrecions d'una regla general: seguint amb l'exemple de la primera llei, el «dany» infligit a un ésser humà pot anar des de crear-li una lleu contrarietat fins a causar-li la mort. Una altra font de problemes és la de regles que entren en conflicte. Establir prioritats entre aquestes regles no acaba de ser una solució: una regla de prioritat més elevada pot tenir una probabilitat mínima d'aplicabilitat i, per tant, no ser rellevant en una situació concreta.

L'ètica basada en el conseqüencialisme topa amb dificultats similars. Wallach i Allen esmenten les quatre habilitats computacionals requerides per un robot conseqüencialista tal com va establir l'informàtic James Gips el 1995:

- 1) Una manera de descriure l'estat del món.
- 2) Una manera de generar possibles accions.
- 3) Una manera de predir la situació o estat resultant del món si s'aplica una determinada acció en l'estat actual.
- 4) Un mètode per a avaluar una situació en termes de bondat o desitjabilitat.

Encara que aquestes habilitats no defineixen certament un algorisme ni un mètode a implementar, sí que es poden considerar com un conjunt essencial d'habilitats que requeriria un robot conseqüencialista. Ara bé, la primera tasca ja planteja la qüestió espinosa de determinar els elements rellevants que defineixen una situació, l'estat del món en aquell moment. Una acció pot afectar molts subjectes, siguin humans, altres éssers vius, fins i tot l'ecosistema o objectes inanimats. On s'ha de detenir l'avaluació de l'estat de cadascun dels potencials afectats? El rang de possibles accions generades (segona tasca) també depèn de la identitat i el nombre de possibles afectats per aquestes accions. Pel que fa a la tercera tasca, si hem estat capaços de solucionar la primera i tenim alguna manera de manejar la incertesa en els efectes associada a l'execució de qualsevol acció (a causa de les contingències que es puguin produir), certament es pot esperar que siguem capaços de predir el resultat *immediat* de l'aplicació d'una determinada acció, però una acció pot tenir també conseqüències a mitjà i llarg terminis, algunes improbables però no impossibles. On

Bibliografia

(Wallach i Allen, 2009, pàg. 87)

s'ha de detenir el càlcul de les possibles ramificacions? Fins a quin punt ens hem d'endinsar en el futur? I, finalment, com s'ha de calcular objectivament la bondat, utilitat o desitjabilitat d'una situació resultant d'una acció? Es poden prioritzar els plaers resultants? És millor l'alegria banal de milers de telespectadors d'un programa porqueria que el plaer estètic d'unes dotzenes d'entesos a escoltar en directe una interpretació musical de qualitat? Es pot quantificar el plaer que experimenta cadascun?

En el fons, les teories que procedeixen de manera descendent des de principis de pretesa universalitat, siguin aquestes formulacions deontològiques del deure o funcions de benefici o utilitat, s'enfronten al mateix problema fonamental quan es planteja la seva implementació eventual en una màquina: delimitar, destriar les dades, fets i circumstàncies rellevants en l'univers de les possibilitats infinites del dia a dia, és a dir, no en una àrea especialitzada i restringida, per a prendre decisions acceptables moralment. Això no és més que una altra versió del *frame problem* del sentit comú i de l'IA general. És el que altres autors denominen **isotropia** (Guarini i Bello, 2014): la rellevància de qualsevol cosa en relació amb qualsevol altra cosa. No obstant això, malgrat ser computacionalment intractables, els principis ètics formulats a alt nivell (això és, com a regles, per exemple) sí que poden constituir heurístiques que guiïn el procés de selecció de la millor acció a emprendre i ajudar en el procés d'avaluació *a posteriori* de les accions fetes pel robot (Allen i Wallach, 2014, pàg. 59).

Una alternativa a la programació explícita d'instruccions és l'aprenentatge. Si gui per demostració o de manera autònoma per reforç, aquests procediments permeten adquirir coneixements i habilitats –en el nostre cas, de discerniment moral– de manera ascendent (*bottom-up*) mitjançant exemples o assajos de prova i error. Altres mètodes que procedeixen de manera ascendent són els que simulen processos evolucionistes, com els algorismes genètics. Aquests procediments, aplicats als AMA, parteixen de la base que la capacitat de discerniment ètic emergirà com un dels aspectes més generals de la intel·ligència. Al cap i a la fi, és per aprenentatge que els humans adquirim les nostres conviccions ètiques (i és possible que la humanitat i les cultures que la integren hagin desenvolupat els codis morals d'una manera anàlogament evolutiva). El problema principal amb els procediments *bottom-up* és determinar el conjunt d'exemples (o de funcions de recompensa o penalització, o els criteris de selecció) que garanteixen l'aprenentatge o l'evolució de comportaments apropiats èticament. S'ha de tenir present que els algorismes d'aprenentatge es limiten a processar dades i que, per tant, els resultats d'aquest aprenentatge poden estar afectats pels biaixos que s'introdueixen involuntàriament i sovint inconscientment en la fase de recopilació de dades. Un exemple d'això són els algorismes experts de concessió de crèdits o els procediments estadístics que atribueixen una major taxa de reincidència delictiva a sectors desfavorits socialment o pertanyents a determinades ètnies: una interpretació acrítica dels resultats perpetua els prejudicis de l'investigador (quan no és conscient que «qui és més

probable que cometi un crim» i «qui és més probable que resulti condemnat» són dues qüestions enterament diferents i que als Estats Units estan molt condicionades per l'ètnia del subjecte). En paraules de Denyse O'Leary (2019):

«Pot un algorisme ser racista? La sortida no té opinió. Com sempre, és l'entrada, que ens ha de preocupar.»

En el cas d'AMA que desenvolupen un criteri ètic d'avaluació per aprenentatge els biaixos inadvertits dels qui seleccionen i anoten les dades amb què s'alimenta el sistema d'aprenentatge són els que conformaran el tipus d'ètica de l'AMA. Potser és preferible que els supervisors de l'aprenentatge sí que introdueixin algun tipus de biaix de manera conscient a fi d'afavorir l'emergència de comportaments ètics desitjables.

«Tots els sistemes d'aprenentatge requereixen biaixos i assumpcions per a funcionar bé, així que no hi perdem res incloent biaixos que són justos èticament.» (Montañez, 2019)

L'ètica de les virtuts es pot considerar com una conciliació dels procediments *top-down* i *bottom-up*: les virtuts poden ser descrites explícitament (i els resultats d'una acció contrastats amb els que resultarien d'un comportament virtuós èticament), mentre que l'adquisició dels trets morals corresponents a un caràcter virtuós es fa per aprenentatge. El sistema cognitiu proposat a Wallach i Allen (2009) per a implementar criteris morals en agents artificials, el model LIDA de Stan Franklin, proporciona mecanismes de selecció d'accions que procedeixen de manera ascendent a partir de col·leccions de dades del sistema sensorial i processos descendents que donen sentit a la situació actual. És discutible que aquest o futurs sistemes que combinin procediments normatius amb aprenentatge basat en l'experiència permetin arribar a obtenir agència moral plena, però és molt possible que permetin aconseguir cotes de moralitat funcional elevades.

2.2. Rellevància ètica dels robots en diferents àmbits de l'experiència humana

Els éssers humans interpretem diferents papers al llarg de la vida tant per evolució vital com dependent de l'entorn on som. A continuació revisarem com pot incidir la presència d'un robot intel·ligent en aquests diferents àmbits, alguns quotidians i uns altres més excepcionals, i les qüestions ètiques que es plantegen a causa d'aquesta presència.

Vegeu també

O'Neil (2016).

2.2.1. L'àmbit laboral: impacte del robot sobre el mercat de treball

La irrupció dels robots en el mercat laboral és percebuda com una amenaça per amplis sectors de la societat. A això contribueixen estudis de projecció amb conclusions que són reinterpretaes i amplificades de manera catastrofista per alguns mitjans de comunicació. Parlarem d'aquests estudis més endavant. De moment analitzem alguns precedents de transformacions significatives en el treball humà, en les seves característiques i possibles analogies amb el cas que ens ocupa. La primera transformació radical del *modus vivendi* humà ha estat possiblement la denominada revolució neolítica, que va començar cap a l'any 10000 aC al Llevant (est del Mediterrani) i als anys 8000-7000 aC a les riberes del Nil, del Creixent Fèrtil (Tigris i Eufrates), de l'Indus i del riu Groc, encara que va tenir lloc durant un llarg període de temps (mil·lennis) en la seva expansió cap a altres llocs habitats per humans. Les societats de caçadors-recol·lectors nòmades es van transformar en societats agrícoles i ramaderes sedentàries.

Des del punt de vista del treball, el més destacat va ser l'**especialització** i, en conseqüència, el comerç. La figura de l'artesà, una persona que després d'anys de formació ha adquirit una sèrie d'habilitats per a confeccionar un tipus de producte determinat, es mantindrà pràcticament fins a la Revolució Industrial i encara existeix als nostres dies (encara que de manera residual). Durant segles els gremis d'artesans mantindran el poder suficient per a oposar-se a la mecanització de treballs que els són propis, tal com il·lustra l'exemple d'una màquina de fer punt que va ser inventada el 1589 per William Lee, a la qual va ser denegada la patent per la Corona anglesa (Frey i Osborne, 2013, pàg. 6). L'emergència d'una classe social capitalista amb creixent influència al Parlament britànic, que al seu torn va guanyar poder enfront de la Corona, va crear el terreny abonat per a la Revolució Industrial que es va produir a la segona meitat del segle XVIII gràcies a la millora de la màquina de vapor aconseguida per James Watt, incloent el regulador de pressió (això és, introduint el control automàtic). La Revolució Industrial va significar la deriva de la demanda de mà d'obra cap al treballador poc qualificat (en comparació de l'artesà) en dividir la manufactura del producte en moltes labors simples i molt especialitzades. Amb l'adveniment de les màquines automatitzades (gràcies als avenços en l'electrònica de control al segle XIX), són precisament aquests treballs de baixa qualificació i molt repetitius que es van veure amenaçats (i eliminats gradualment). Hi ha una nova demanda de treballadors amb certa qualificació tècnica, especialistes en electromecànica i disciplines afins, capaços de manejar i mantenir maquinària cada vegada més complexa. Així, doncs, assistim de nou solament a un desplaçament de la demanda de mà d'obra, circumscrit al sector secundari una vegada més.

No obstant això, amb la flexibilitat i la sofisticació creixent que introdueix la informatització de les màquines, i en particular dels robots, l'espectre de treballs en què els humans poden veure's desplaçats per màquines es va ampli-

Reflexió

Hi ha evidència de llavors plantades en dates fins i tot més antigues, però no comporten els canvis d'hàbits associats al sedentarisme.

ant. Les revolucions tecnològico-industrials anteriors han causat l'eliminació de certs tipus de treball però n'han creat d'altres, i han tingut lloc a un ritme en què la transició, l'adaptació dels treballadors als nous requisits, s'ha pogut fer de manera gradual (fins i tot pel que fa al relleu intergeneracional). Que passi el mateix en el present cas és qüestionable tant per la rapidesa dels canvis com per l'abast amb què poden tenir lloc. Els ordinadors i els robots tenen el potencial de substituir la força de treball humana en els tres sectors d'una manera sense precedents en la història de la humanitat.

En primer lloc analitzem el sector industrial, on ja fa dècades que s'han anat introduint robots. Els robots han anat substituint treballadors humans en tasques perilloses (com alimentar premses o cisalles), en ambients nocius (per als quals es necessiten proteccions especials, com per exemple en pintura amb presència de dissolvents), en labors monòtones (tasques repetitives com en una cadena de muntatge) o en combinacions de les anteriors. Malgrat substituir les persones que feien aquestes labors fins al moment, hi ha un cert consens en què els robots han comportat una millora de les condicions laborals en general. La introducció de robots ha comportat crear nous llocs de treball, sigui directament a les mateixes plantes com a personal per a programar i mantenir els robots, sigui indirectament com a personal empleat a les fàbriques de robots, personal de formació i entrenament, i personal per a recerca i desenvolupament. Un estudi diu:

«Entre el 2009 i el 2014 les 62 companyies amb major base de robots instal·lada van crear 1,25 M de nous llocs de treball nets, el 20 % més.» (Robotonomics, 2015)

Es mantindrà aquesta tendència o arribarà el moment en què altres màquines seran capaces de programar robots per a tasques concretes, apareixeran robots de manteniment o la demanda de robots industrials i, per tant, la indústria de fabricació de robots arribarà a una certa saturació? La veritat és que la indústria aposta pel que pot reportar el màxim benefici en el termini menor, i sol ser bastant conservadora pel que fa a introduir tecnologies que no han estat prou provades. Per aquest motiu, encara que la substitució de treballadors humans per màquines resulti atractiva a l'empresa per la reducció de costos associada, ho és més la introducció d'equips mixtos humans-robots capaços de treure partit de les millors qualitats de cadascun. Això és possible gràcies al desenvolupament dels **cobots** (*cooperative robots*), robots que poden treballar en col·laboració i contacte amb l'ésser humà sense comportar un risc per a la seva integritat física. Si es confirma aquesta tendència, la destrucció de llocs de treball en la indústria seria molt inferior a l'augurada per les prediccions més pessimistes. Així i tot, un informe recent d'Oxford Economics (2019) xifra en vint milions els llocs de treball que es perdran en la indústria manufacturera d'aquí al 2030. A tall de comparació, Nacions Unides xifra la pèrdua de llocs de treball per l'estrès tèrmic a causa del canvi climàtic en vuitanta milions el 2030 (Notícies ONU, 2019), majoritàriament en el sector agrícola i a països

de l'Àsia meridional i l'Àfrica occidental (encara que en aquest cas es tracta de l'equivalent a la suma d'hores de treball perdudes per excés de calor més aviat que llocs de treball que desapareixen pròpiament).

Pel que es fa a altres sectors, la mecanització també ha comportat un desplaçament d'un gran nombre de treballadors en el passat. Les recol·lectores automàtiques, per exemple, van comportar una reducció dràstica de la mà d'obra requerida en tasques associades a la sega. Els caixers automàtics van tenir un impacte igualment notable en la banca, i més recentment les compres i reserves en línia han afectat el minorista i les agències de viatges, per anomenar alguns exemples. Quina serà la incidència de la computerització i robotització del sector primari i el de serveis? En un estudi del 2013 de l'Oxford Martin School i del Departament de Ciències de l'Enginyeria de la Universitat d'Oxford, els colls d'ampolla en la susceptibilitat d'un treball a ser computeritzat són les habilitats perceptives i la destresa manual, la creativitat, i la intel·ligència social (Frey i Osborne, 2013, pàg. 24 i seg.).

Els treballs que exigeixen un major grau en qualsevol d'aquests factors són els menys susceptibles de ser computeritzats. Per exemple, encara que la labor de venda a distància requereixi una certa interacció amb el client, no són necessàries les habilitats socials d'un conseller matrimonial, que ha de ser capaç de mostrar un alt grau d'empatia. L'informe suggereix que es produiran dues onades en la computerització de treballs, la primera deguda a tecnologia d'aprenentatge automàtic i de sensors ja existent, que afectarà labors en el sector de transports i logística, labors de suport administratiu i labors de producció, i la segona amb requisits de superar els colls d'ampolla tecnològics esmentats, de manera que els treballs que exigeixen creativitat i un grau d'intel·ligència social elevat se situaran en l'escala dels de menor risc de ser computeritzats. Aquesta conclusió és la mateixa a què arriba l'estudi més recent esmentat abans:

«Serà difícil que les màquines reemplacin els humans en ocupacions del sector de serveis que requereixin compassió, creativitat i intel·ligència social.» (Oxford Economics, 2019, pàg. 7)

Aquest segon estudi constata el desplaçament de treballadors del sector productiu, per l'automatització o robotització, cap als sectors del transport, construcció i manteniment, i treball administratiu i d'oficina, que són justament els sectors més vulnerables a l'automatització durant les properes dècades. Malgrat això, els autors desaconsellen dur a terme polítiques per a desaccelerar l'adopció de tecnologia robòtica, ja que aquesta repercuteix globalment de manera molt positiva en l'economia, i suggereixen, en canvi, que el *dividend robòtic* es destini en part a mitigar la situació dels treballadors més vulnerables i a afavorir-ne l'adaptació i integració en el nou panorama laboral i social.

Habilitats de l'ésser computeritzat

Cadascun d'aquests colls d'ampolla és desglossat al seu torn en altres factors; per exemple, en la intel·ligència social es considera la perceptibilitat social, la capacitat de negociació, la persuasió, i l'assistència i cura dels altres.

Reflexió

És important destacar que l'informe insisteix que en l'estudi considera únicament treballs existents actualment, i no especula sobre possibles ocupacions que emergeixin precisament per motiu de la computerització i la robotització.

2.2.2. L'àmbit del transport: vehicles autònoms

Un vehicle autònom és un mitjà de transport automotriu i dotat de sensors que li permeten percebre el seu entorn i arribar a la destinació sense la intervenció d'un conductor humà. Un avió que opera en mode de pilot automàtic es pot considerar un vehicle autònom, però el terme es refereix en la seva accepció més difosa a vehicles terrestres, en concret cotxes i camions. Els sensors i sistemes requerits per a la conducció autònoma inclouen detectors de proximitat ultrasònics, radar i lidar, càmeres de visió i GPS. També es requereix un ordinador a bord capaç de processar tota la informació subministrada pels sensors i interpretar-la en termes de vies de circulació, senyals de trànsit, altres vehicles, vianants, ciclistes i, en general, obstacles fixos i mòbils a la calçada. La possibilitat de comunicació amb altres vehicles autònoms i/o amb sistemes centralitzats de gestió de trànsit obre les vies per a una circulació més segura i eficient, amb menys accidents i més fluïdesa. La incidència del vehicle autònom es produeix tant en l'àmbit laboral (transport de mercaderies i de persones) com en el privat.

L'Estàndard SAE J3016 defineix els sis nivells d'autonomia següents (SAE International, 2019):

- Nivell 0. No hi ha autonomia (pot haver-hi avisos i control puntal).
- Nivell 1 (*hands-on*). Hi ha control compartit. Per exemple: Adaptive Cruise Control (l'humà atén a la direcció del vehicle mentre el sistema controla la velocitat), Parking Assistance (viceversa). L'humà ha de poder accedir al control total a qualsevol moment.
- Nivell 2 (*hands-off*). El sistema controla la direcció, l'acceleració i la frenada. L'humà supervisa i està disposat a intervenir en qualsevol moment.
- Nivell 3 (*eyes-off*). L'humà pot distreure's, llegir, mirar una pel·lícula..., però ha de poder intervenir en un temps determinat, avisat pel sistema.
- Nivell 4 (*mind-off*). L'humà pot dormir o abandonar el seient del conductor en àrees o situacions determinades. Fora d'aquestes, si l'humà no reprèn el control, el sistema avorta el viatge (aparca).
- Nivell 5 (*steering wheel optional*). No hi ha cap intervenció humana (per exemple, un taxi robot).

Cada nivell defineix un repartiment diferent de responsabilitats entre el conductor i els fabricants o programadors del sistema en cas d'accident. Entre els avantatges del vehicle autònom hi ha:

- Disminució del risc d'accidents, ja que s'eliminen possibles imprudències (consum d'estupefaents, saltar-se les regles o senyals de trànsit, distraccions, etc.) i s'eviten altres vehicles gràcies a una xarxa d'interconnexió.
- Major comoditat per a usuaris, millora de la mobilitat per a no conductors.
- Reducció d'embotellaments de trànsit, pacificació del trànsit urbà i interurbà.
- Afavoriment del vehicle compartit, menor ocupació de la via.
- Reducció dels costos de les assegurances.
- Major eficiència energètica.

En contra, es poden esgrimir els arguments següents:

- Resistència dels conductors a abandonar els hàbits.
- Augment de la desocupació en els sectors del transport i de serveis associat (restauració, hostaleria...).
- Privadesa compromesa (seguiment dels desplaçaments).
- Desplaçament de la població a àrees suburbanes, major invasió humana del territori.
- Possibilitats que ofereix per a pirateig, control remot, terrorisme.

Des del punt de vista ètic, una de les qüestions més rellevants i polèmiques és el de les reaccions programades a situacions límit. Són variants del que es coneix en ètica com el **dilema del tramvia**:

«Un tramvia corre fora de control per una via. En el seu camí hi ha cinc persones lligades a la via per un filòsof malvat. Afortunadament, és possible accionar un botó que encaminarà el tramvia per una via diferent. Per desgràcia, hi ha una altra persona lligada a aquesta via. S'hauria de prémer el botó?» (Foot, 1978)

Una versió d'aquest dilema, ja en el context de la conducció d'un vehicle, és el **dilema del túnel**:

«Viatges sol en un vehicle autònom per una carretera de muntanya d'un sol sentit i t'apropes ràpidament a un túnel estret. Just abans d'entrar al túnel un nen intenta creuar corrent, però ensopega i cau de manera que bloqueja l'entrada al túnel. El vehicle [no pot frenar prou ràpid per a evitar colpejar el nen, per la qual cosa] té dues opcions: colpejar i matar el nen o girar el volant i xocar contra la paret del túnel i matar-te a tu. Com hauria de reaccionar el cotxe?» (Miler, 2014)

Una pregunta relacionada és: qui hauria de prendre aquesta decisió? L'autor del dilema i de l'article sosté que no hauria de ser una decisió imposada per disseny, ja que els constructors o programadors del vehicle estarien incorrent d'aquesta manera en *paternalisme* pel que fa a decisions que competeixen el conductor afectat directament. En una enquesta duta a terme en el mateix portal web (Open Roboethics Initiative, 2014) el 64 % dels enquestats van declarar que si ells anessin com a passatgers del vehicle aquest hauria de seguir recte i atropellar el nen. Pel que fa a la pregunta de qui hauria de determinar la resposta del vehicle, el 44 % va apostar pel passatger, el 33 % pels legisladors, el 12 % pel fabricant o dissenyador i l'11 % per altres.

En una plataforma d'internet muntada per investigadors del MIT és possible jutjar diferents variants del problema del cotxe autònom que ha de decidir entre atropellar vianants que creuen un carrer o estavellar-se contra una barriera sòlida (o desviar-se i atropellar un altre grup de vianants). Tant els ocupants del vehicle com els vianants varien en nombre i característiques (edat, gènere, estat de salut, reconeixement social, etc., inclosos fins i tot animals de companyia), i també hi ha variants com si els vianants creuen en vermell o en verd, o diferents graus d'afectació per l'accident (mort, ferides de diferent gravetat, etc.). Els resultats obtinguts fins a mitjan 2018 entre milions d'enquestats de 233 països i territoris estan recollits i analitzats en un article de la revista *Nature* (Awad *et al.*, 2018). Les diferències culturals queden en evidència en l'estudi: per exemple, països pròspers amb institucions fortes són menys propensos a salvar vianants que han creuat en vermell. Això il·lustra les dificultats d'elaborar un codi ètic vàlid universalment, encara que la finalitat de l'estudi no era aquesta sinó remarcar aquestes diferències culturals. El més probable és que els cotxes es limitin a seguir les normes de circulació de cada país (que, de fet, ja preveuen algunes contingències més probables, com que a prop d'un parc infantil o un col·legi pugui sortir inesperadament un infant entre els cotxes, per la qual cosa cal circular amb precaució especial en aquestes zones), i tindran una assegurança per a cobrir danys.

MIT

<http://moralmachine.mit.edu/>

Una altra qüestió relacionada és que una reacció preprogramada es pot interpretar com un algorisme que *apunta* a determinats col·lectius o individus. Citant l'exemple de Miler (2017), un sistema que busqui minimitzar el dany causat «preferirà» atropellar un ciclista que porti casc (i compleixi així les normes de tràfic) que un que no en porti. En aquest cas estaria penalitzant els ciclistes responsables.

2.2.3. L'àmbit del conflicte: robots militars

En un conflicte armat la tipologia de robots comprèn les categories següents segons la missió que duguin a terme:

- Missions humanitàries i tàctiques no cruentes: desminatge d'una àrea, desactivació d'explosius, rescat de ferits, reconeixement, logística i transport.
- Missions tàctiques defensives: dispositius muntats sobre tancs, vaixells o altres vehicles, o bé situats en una frontera o perímetre a defensar (sistemes antimíssils), que responen de manera immediata i automàtica a qualsevol atac detectat en altura.
- Missions tàctiques ofensives: armes mòbils (majoritàriament per aire, encara que també n'hi ha que es desplacen per terra) emprades en la neutralització de l'enemic.

Des del punt de vista ètic, les LAW (*lethal autonomous weapons*, armes letals autònomes) tenen una significació especial. Hi ha diferents interpretacions pel que fa al grau d'autonomia: alguns autors la formulen a molt alt nivell, entenent com a tal que la màquina descriu un objectiu global i fixa els objectius particulars, els cerca i detecta, i els elimina. Per a altres autors, el qualificatiu d'autònom ja és aplicable si el robot és capaç de detectar i seguir un objectiu concret, encara que la decisió de fer foc recaigui en un ésser humà. El format més corrent en l'actualitat és la d'un vehicle aeri no tripulat (UAV, *unmanned aerial vehicle*) equipat amb míssils que poden ser disparats sobre l'objectiu. Aquests UAV disposen de càmeres que poden proporcionar un senyal de vídeo a l'operador humà, de manera que aquest estarà a l'aguait de la situació en tot moment. L'operador humà pot estar «en el bucle» (*human in the loop*), teleoperant el vehicle la major part del temps excepte durant la fase més rutinària de vol en pilot automàtic, o bé «sobre el bucle» (*human on the loop*), on es limita a monitorar el funcionament del vehicle mentre aquest s'encarrega de manera autònoma de les tasques d'enlairament i aterratge, i de la determinació i cerca de l'objectiu. Algunes operacions són dutes a terme per un grup (*swarm*, eixam) d'UAV de manera coordinada: davant la impossibilitat de controlar cada vehicle de manera individual, s'imposa la manera d'operar *on-the-loop*. En qualsevol dels casos, hi ha el consens generalitzat que ha de ser un humà qui prengui la decisió final sobre establir o eliminar un objectiu específic. D'aquesta manera, l'atribució de responsabilitats per les accions o fins i tot pel mal funcionament del robot se situa clarament en els elements humans de la cadena de comandament. Això és evident en el cas dels robots sense capacitat deliberativa sobre les seves accions. Tornarem a tractar el tema de la responsabilitat i els robots autònoms en la tercera part.

Es pot dir que aquests sistemes constitueixen la culminació d'un procés observat al llarg de la història de la humanitat, i més específicament dels conflictes bèl·lics, consistent a anar augmentant la distància entre els contendents per contrarestar les dues pors a què s'enfronta el soldat: a morir i a matar (Sharkey, 2014). En un UAS l'enemic pot ser a milers de quilòmetres de distància de l'operador. A la distància física se suma l'efecte de seguir l'operació amb una pantalla, que s'assembla a la dinàmica dels videojocs, la qual cosa incrementa el factor de desaferrament respecte de la realitat de l'escenari bèl·lic.

LAW

També és coneguda com a LAWS (*lethal autonomous weapon system*), LAR (*lethal autonomous robot*), *robotic weapon* (arma robòtica) o, més popularment, *killer robot* (robot assassí).

UAV

També és conegut per l'anglicisme *dron*. El terme anàleg UAS (*unmanned aerial system*) es refereix al conjunt format pel vehicle aeri i l'estació de control en terra, el sistema de comunicacions, etc.

Human out of the loop

També podria estar «fora del bucle» (*human out of the loop*) amb el sistema en operació de manera autònoma, sense supervisió.

El dret internacional humanitari (DIH) estableix els principis següents relatius a les accions militars durant un conflicte armat (*ius in bello*):

- **Necessitat militar.** Un atac o acció ha de servir per a contribuir a derrotar l'enemic. Ha de ser un atac sobre un objectiu militar legítim.
- **Proporcionalitat.** Els danys causats a civils o propietats civils han de ser proporcionals i no excessius en relació amb els beneficis militars previstos.
- **Distinció.** Els combatents han de ser capaços de distingir combatents de civils. Les lleis de protecció s'estenen fins i tot a combatents ferits o que s'han rendit.

També hi ha un consens generalitzat a no utilitzar armes o mètodes que generin un sofriment superflu, gratuït i indiscriminat, com queda reflectit en disposicions addicionals del DIH que s'esmenten més endavant.

Un robot combatent totalment autònom complirà millor o pitjor els principis del DIH? Alguns autors sostenen que l'absència d'emocions com la por o la sed de venjança facilita actuar amb objectivitat i ajustar el propi comportament als preceptes del DIH (Arkin, 2009). Lokhorst i Van den Hoven (2014) consideren que un robot bèl·lic pot ser programat per a la incapacitació temporal de l'enemic abans que matar-lo, i procedir d'aquesta manera fins i tot en situacions en què un soldat humà no té possiblement altra opció que matar. Ara bé, hi ha reserves fonamentades sobre el fet que els sistemes de percepció actuals (i del futur proper) amb què estan equipats els robots permeten interpretar una escena de manera que es garanteixin els principis de proporcionalitat i distinció. No és solament una qüestió sensorial, sinó d'interpretació: pel que fa a la proporcionalitat, «no hi ha cap mètrica coneguda per a mesurar objectivament el sofriment innecessari, superflu o desproporcionat» (Sharkey, 2014, pàg. 123). I, quant al principi de distinció o de discriminació, requereix sovint entendre el context i les intencions de l'altre³.

En la guerra moderna no hi ha una distinció manifesta basada en l'uniforme: moltes vegades es tracta dels denominats «insurgents», que no vesteixen de manera especial. Abunden les situacions que requereixen una capacitat de discriminació fina i matisada: nens obligats a portar fusells descarregats, insurgents que enterren els seus morts o l'exemple esmentat a Guarini i Bello (2014, pàg. 130) d'un parell de nens *sikhs* que juguen a perseguir-se amb el *kirpan* (punyal ritual) en el moment que les forces de contrainsurgència entren a la casa on són. Aquests autors insisteixen en el paper fonamental que les emocions tenen per a resoldre de manera eficient el problema associat a la isotropia i la teoria de la ment, en què un procés deliberatiu necessitaria un temps de què no es disposa i/o uns recursos computacionals extraordinaris.

Dret internacional humanitari

El DIH s'aplica únicament en cas de conflicte armat. Està contingut en els quatre convenis de Ginebra de 1949, dels quals són signants gairebé tots els estats. Aquests convenis es van completar amb els protocols addicionals de 1977 relatius a la protecció de les víctimes dels conflictes armats. (Extret del document del Comitè Internacional de la Creu Roja *Què és el dret internacional humanitari?* <https://www.icrc.org/es/doc/assets/files/other/dih.es.pdf>)

⁽³⁾ El que en filosofia i psicologia s'entén com a teoria de la ment, la capacitat d'atribuir estats de la ment als altres, això és, creences, intencions, desitjos, emocions, etc.

Reflexió

El fet que un sistema artificial pugui simular quelcom que compleix la comesa computacional de les emocions no significa que aquest sistema senti realment, com també indiquen els autors.

Un estat pot apostar per la tecnologia robòtica per reduir els efectius humans en el teatre d'operacions i, per tant, les possibles baixes humanes (entre les seves files). D'altra banda, aquest mateix factor pot facilitar que aquest estat tingui menys reticències a l'hora d'emprendre accions bel·ligerants en detriment de la diplomàcia, ja que no haurà d'enfrontar-se a l'efecte impopular del «retorn de taüts». Encara més, el factor econòmic també serà molt rellevant. Tal com s'apunta a Kahn (2017), no solament el cost associat a la pèrdua de vides humanes (que també es tradueix en pensions, indemnitzacions, etc.) té repercussions, sinó que un UAV capaç de carregar una bomba de 250 kg com el MQ-9 Reaper costa aproximadament una quarta part del que costa un avió de reacció amb la mateixa capacitat tripulat, com l'F/A-18F, i el cost associat a entrenar un operador d'UAV està en la relació d'1 a 360.000 respecte al d'entrenar a un pilot. L'efecte net pot ser l'augment de la conflictivitat, que també produirà més víctimes humanes malgrat la presència de robots en les conteses. A això cal sumar que els estats en desavantatge tecnològic optin per altres vies com el terrorisme.

El 2013 es va iniciar la campanya internacional «Stop killer robots»: 272 científics, enginyers, experts en IA i robòtica de trenta-set països van demanar prohibir el desenvolupament i instal·lació d'armes plenament autònomes. La qüestió és si prohibir les LAW garantirà que no s'utilitzin. Disposicions addicionals del DIH fan referència a la prohibició d'armes específiques, com la Convenció sobre Armes Biològiques (1972), la Convenció sobre Certes Armes Convencionals i els seus cinc protocols (1980, es tracta de restriccions sobre armes de fragmentació, armes trampa, armes incendiàries o armes làser engegadores), la Convenció sobre Armes Químiques (1993) i el tractat d'Ottawa de 1997 sobre les Mines Antipersona. A més, el 1970 es va signar el tractat de No-prolifерació Nuclear⁴, a revisar el 2020, i el 2017 el tractat sobre la Prohibició de les Armes Nuclears (que ha estat ratificat solament per onze països (a juliol de 2018)⁵.

Així i tot, els Estats Units i Rússia tenen prop de quinze mil armes nuclears (en total). En molts casos la prohibició d'una arma arriba anys després d'haver estat utilitzada, i moltes vegades no tots els països signen o ratifiquen els acords. En països com els Estats Units tota nova arma requereix una autorització en què s'ha de demostrar que es pot usar complint el principi de proporcionalitat, que pot ser controlada de manera que solament vagi dirigida a un objectiu específic acceptable legalment i que no hi ha cap tractat o llei nacional que la prohibeixi. No obstant això, la combinació de dispositius i armes ja aprovats no ha de passar necessàriament l'examen, tal com va determinar l'oficina del cos jurídic de les forces armades americanes (Judge Advocate General) en el cas d'un UAV, el Predator UCAV, destinat inicialment a vigilància i que va ser armat amb míssils Hellfire (Sharkey, 2014, pàg. 119).

⁽⁴⁾ Amb l'excepció d'Índia, Pakistan, Israel, Sudan del Sud i Corea del Nord.

⁽⁵⁾ Es requereix la ratificació d'almenys cinquanta països perquè entri en vigor.

Stop killer robots

A dia d'avui (octubre de 2019) la coalició inclou 118 ONG de 59 països.

I què passa quan no són les forces armades regulars d'un país democràtic les que utilitzen aquest tipus d'armament? Un militar de carrera concedirà generalment gran importància a l'honor, que en el seu cas va associat a l'acompliment de l'activitat respectant les «regles del joc»; és el que es denomina «ètica del soldat» (O'Meara, 2014). A més, és conscient que les seves accions seran sotmeses a escrutini dins o fora de l'àmbit militar. No passa el mateix amb altres col·lectius. Les agències d'intel·ligència, per exemple, procedeixen de manera opaca: la CIA ha fet moltes accions usant UAV per a «decapitar» cèl·lules insurgents, emparant-se en el fet que es tracta d'operacions encobertes per a no donar compte d'aquestes execucions extrajudicials (Sharkey, 2014, pàg. 114). A aquestes persones se'ls ha despullat, a més de la vida, del dret al recurs o a la rendició. Paramilitars, mercenaris que actuen com a forces de seguretat de grans empreses i terroristes són altres col·lectius que actuen al marge dels acords internacionals de la *ius in bello*.

2.2.4. L'àmbit de la salut: pròtesis, sistemes experts i robots cirurgians

Sigui per accident o agressió (això és, per trauma), per malaltia adquirida o per malformació genètica, els éssers humans estem sotmesos a l'eventualitat d'una pèrdua funcional sensorial o motriu associada a la deterioració o eliminació d'un òrgan, extremitat o teixit neuronal. Al llarg de la història hem intentat suplir alguna d'aquestes pèrdues amb accessoris o artefactes, i és paradigmàtica la imatge del pirata amb cama de pal, garfi i ull de vidre o pegat (que en aquest cas compleix únicament una funció estètica, encara que tampoc no se n'ha de menystenir la importància). Però no hi ha dubte que és en temps actuals i anys esdevenidors que els avenços tecnològics permetran disposar de pròtesis cada vegada més perfeccionades i funcionals. Una part dels avenços provenen de la ciència dels materials, que permeten obtenir pròtesis cada vegada més lleugeres i resistents, i més compatibles amb els teixits humans. Però és sens dubte mitjançant la robòtica que s'han aconseguit els majors assoliments quant a recuperació de la mobilitat perduda. El paradigma *percep-pensa-actua* té ple sentit en aquest cas: els biosensors capten l'activitat muscular o neuronal dels teixits funcionals del pacient, el controlador s'encarrega de processar aquests senyals i els que provenen de la mateixa pròtesi (per exemple, posició dels dits o força d'impacte al peu durant la trepitjada), i els actuadors són els motors o dispositius que generen el moviment o l'adaptació activa de la pròtesi. La funció cognitiva queda repartida d'aquesta manera entre el controlador i el mateix cervell del pacient.

En principi, no hauria d'existir cap objecció ètica a les pròtesis que supleixen una funcionalitat perduda: es tracta de restaurar les capacitats que un ésser humà ja tenia o que són comunes en un ésser humà. No obstant això, el debat ètic està justificat en almenys tres qüestions relacionades amb aquesta tecnologia:

- **Seguretat.** Aquesta qüestió inclou dos aspectes: fiabilitat i intrusió. La fiabilitat es refereix al funcionament correcte del producte, que compleixi no solament les expectatives de l'usuari sinó les especificacions del fabricant. En cas d'un producte amb les característiques d'una pròtesi, aquest aspecte és especialment sensible, ja que no és un mer aparell que deixa de funcionar (o funciona malament) i l'usuari pot seguir la seva vida mentre és reparat, sinó que és la mateixa integritat i mobilitat o funcionalitat de l'usuari que es posa en una situació difícil. Un problema que en altres circumstàncies no passa de ser una molèstia, com és el sobreescalfament de les bateries, en el cas d'una pròtesi pot causar cremades a l'usuari (a més d'espatllar la pròtesi). Pot donar-se el cas que una pròtesi causi dany a un tercer (per exemple, una mà robotitzada que estreny tant la mà d'una persona a la qual saluda que li causa lesions). Tant en l'un cas com en l'altre caldrà determinar si l'origen de l'error està en un mal disseny, una fabricació defectuosa o un mal ús de la pròtesi. El segon aspecte, el de la intrusió per part d'un *hacker* que pren el control sobre la pròtesi, és especialment tenebrós, ja que obre la porta a un nou tipus de criminalitat. Ja s'han donat casos de *hackers* que havien pogut accedir a controlar bombes d'insulina de pacients diabètics, als quals haurien pogut assassinar amb una simple sobredosi d'insulina. En aquest cas la responsabilitat recau evidentment sobre el ciberintrús, encara que també pot ser que el fabricant no hagi inclòs prou tallafocs per a evitar-ne l'accés o que l'usuari hagi estat negligent a l'hora de posar els corresponents antivirus al dia.
- **Desigualtat.** Les pròtesis robòtiques tan avançades al principi estaran solament a l'abast d'una minoria amb prou poder econòmic. Fins i tot entre individus amb poder adquisitiu insuficient pot donar-se algun tipus de discriminació, com en el cas d'un estat que finança la pròtesi d'un soldat que ha anat a lluitar voluntàriament en una guerra llunyana i, en canvi, no la d'algú que ha perdut un membre en un accident. Una pròtesi no serà mai un producte de consum de masses, no solament perquè els amputats representen afortunadament un percentatge reduït de la població total, sinó sobretot perquè cada pròtesi ha de ser adaptada a un usuari únic, i al cost de la pròtesi mateixa cal afegir el de la rehabilitació i entrenament del pacient. Amb l'avenç i la millora de les pròtesis, aquestes poden convertir-se en un altre factor de desigualtat social més acusat, tal com es comenta a continuació.
- **Capacitats sobrehumanes.** Ja en l'actualitat una pròtesi, juntament amb les seves limitacions i defectes, també aporta una sèrie d'avantatges a l'usuari: l'absència de dolor experimentat⁶ i la duresa dels materials constitutius poden atorgar un cert avantatge si l'usuari decideix emprar-la agressivament. Més enllà de restaurar capacitats, en el futur una pròtesi pot millorar aquestes capacitats per sobre de les possibilitats biològiques: pot atorgar major força, velocitat, resistència, destresa... Pròtesis cerebrals podrien millorar capacitats cognitives com l'aprenentatge o la memòria, i pròtesis sensorials obrir tot un nou món de percepcions. Aquestes perspec-

⁽⁶⁾ No estem parlant de possibles nafres que poden produir-se per fregament en la connexió amb el membre romanent, que poden ser molt doloroses.

tives poden induir (a qui s'ho pugui permetre) a prescindir de membres i òrgans sans per substituir-los per les versions artificials millorades. És ben cert que els humans pretenem evolucionar i millorar com a espècie, però aquesta tendència crearia sens dubte una elit a curt termini que afegiria a la seva fortalesa financera un avantatge deslleial en termes físics i psicològics. La complexitat de les causes legals que es derivarien dels conflictes inevitables entre humans biològics i humans millorats seria certament elevada. D'entrada, les recomanacions presentades al Parlament Europeu el 2017 promouen la creació de comitès d'ètica als hospitals que incloguin experts en roboètica per fer front a les decisions sobre la conveniència d'implantar pròtesis per a millores específiques.

Sense abandonar el camp de la salut humana, altres elements de gran significació ètica són els sistemes experts utilitzats per a la diagnòsi mèdica i els robots cirurgians. Els primers són programes d'IA que a partir de la simptomatologia (en forma de preguntes a les quals les anàlisis i proves mèdiques van donant resposta) són capaços de proporcionar possibles diagnòstics. Aquests sistemes experts poden contribuir en gran manera a reduir l'error mèdic. Com sempre, el dilema ètic emergeix quan les coses van malament: qui és responsable d'un diagnòstic erroni, que possiblement ha portat a un tractament inadequat amb conseqüències nefastes per al pacient? ¿Els experts mèdics que van participar en el desenvolupament del sistema, els programadors, el metge que l'ha utilitzat, el personal que ha fet les proves o ha introduït les dades? En principi, una recerca hauria de poder dilucidar on rau la font de l'error. També podria donar-se el cas d'un metge que creu que el diagnòstic que proporciona el sistema expert és erroni, però decideix donar-lo per bo per no carregar amb el risc d'equivocar-se.

Pel que fa als robots cirurgians, a dia d'avui actuen generalment com a braços telemanipuladors en el sentit que no operen de manera autònoma sinó reproduint els moviments del cirurgià, encara que amb un pols molt més estable i a escala microscòpica. Per regla general, la seva introducció ha comportat una millora en la qualitat de les intervencions, que resulten molt menys invasives i, per tant, redueixen el postoperatori i afavoreixen la recuperació del pacient. Possiblement, han evitat un nombre considerable de defuncions a la taula d'operacions, però no són del tot infalibles, i també s'ha donat el cas d'una defunció durant una operació per una fallada en el sistema de control. Una fallada d'aquest tipus pot ser de disseny, de fabricació o de programació, i més difícilment d'ús, ja que precisament aquests sistemes incorporen salvaguardes per a evitar reproduir moviments bruscos o desmesurats del cirurgià. No obstant això, també pot passar que un cirurgià hagi observat alguna anomalia en el sistema i, malgrat això, hagi decidit seguir endavant amb l'operació. Si això es pogués demostrar, una part de la responsabilitat recauria sens dubte en el mateix cirurgià, encara que aquest podria ser víctima d'un excés de confiança (*overtrust* en anglès) en el sistema robòtic pel mer fet de ser un sistema automàtic. Es tracta d'un fenomen transversal: el metge al qual al·ludíem al final del paràgraf anterior podria ser un altre cas de confiança excessiva, com

la d'uns pares (o un fisioterapeuta) que tenen una seguretat desmesurada en la pauta imposada per un robot rehabilitador a pesar que després de les primeres repeticions el pacient infantil mostra signes de malestar evidents (Borenstein *et al.*, 2017). Aquesta confiança excessiva o fins i tot cega en els sistemes automàtics també es dona en altres àmbits, com per exemple el bèl·lic, amb morts causades per foc amic en un sistema automàtic encara que hi hagi un humà en el bucle (Kirkpatrick *et al.*, 2017), el del transport, amb un accident d'aviació causat per confiar en lectures errònies del pilot automàtic, o en els conductors que van a un llac, el mar, un barranc o un desviament de més de 2.500 km amb el cotxe confiant en el GPS (Borenstein *et al.*, 2017). Els mateixos sistemes han de ser dissenyats (o els usuaris entrenats) probablement per a no generar aquest tipus d'excés de confiança.

2.2.5. L'àmbit domèstic: seguretat, privadesa, respecte a la intimitat

La frase del segle XVI «la casa d'un home és el seu castell» posa en relleu la distinció, espacial en aquest cas, entre l'esfera pública i la privada. L'exterior és sinònim d'estar exposat als ulls de la societat, l'Estat, els veïns i transeünts, i també a innombrables perills, que s'assumeixen en traspasar el llindar. Portes endins és l'àmbit de la casa, la privacitat, la intimitat i també la seguretat. De manera metafòrica, els humans també delimitem i separem ambdues esferes i vigilem gelosament que no es traspassi la frontera entre elles. Així i tot, avui dia la privadesa està més en dubte que mai. Això passa de manera molt inadvertida: mitjançant l'ús que fem d'ordinadors personals i telefonia mòbil, programes espia, no necessàriament de delinqüents cibernètics ocults sinó d'empreses o institucions, poden seguir els nostres moviments i les nostres transaccions, i també poden elaborar un perfil de les nostres preferències i afinitats mitjançant l'anàlisi de les dades massives, la ingent quantitat de dades que cada dia aboquem a internet. No sempre passa de manera inadvertida: som sovint nosaltres mateixos que a les xarxes socials exposem les nostres dades, predileccions i febleses en un exercici d'exhibicionisme digital propi d'una societat narcisista i egocèntrica. Malgrat això, creiem retenir un cert control sobre el que volem mostrar, i en el pitjor dels casos el que exposem són dades i el que pot resultar danyat és la pròpia imatge (encara que a vegades el dany també pot ser econòmic). Un cas a part és el dels menors (i el d'alguns adults inconscients), que poden veure's exposats a alguna mena de ciberassetjament. El panorama pot donar un nou salt qualitatiu amb la introducció dels robots, capaços de percebre, processar, guardar en memòria i compartir tot el que passa al seu voltant, i de desplaçar-se d'un lloc a un altre.

Ryan Calo (2014) distingeix tres aspectes en què els robots posen la privadesa en dubte:

- Faciliten la **vigilància directa**, ja que poden revestir qualsevol forma i grandària, no necessiten descans ni tenen distraccions, i poden estar equipats amb molts sensors que excedeixen les capacitats perceptuals huma-

nes. No solament els governs poden abusar d'aquestes noves capacitats de vigilància, sinó corporacions, mitjans i individus en

«àmbits tan diversos com seguretat, *voyeurisme* o màrqueting» (Calo, 2014, pàg. 188).

- Proporcionen noves **vies d'accés** a espais protegits tradicionalment, especialment el domicili particular, mitjançant el robot domèstic. El *hacker* pot accedir a fitxers i documents, i també a habitacions, objectes i persones. El debat legal qüestiona, entre altres coses, que hi hagi hagut una invasió punible de la privadesa prèviament al fet que algun ésser humà hagi accedit a la informació recollida pel robot. Si l'usuari accedeix que el seu robot domèstic o personal aboqui la informació recollida en un servidor extern per a ser processada o que estigui connectat a internet no solament per al seguiment per part del mateix usuari sinó per a actualitzacions o cerca d'informació per a millorar les seves prestacions, està acceptant comprometre d'alguna manera la seva pròpia privadesa, la qual cosa pot comportar al seu torn una reducció voluntària dels drets constitucionals sobre aquesta privadesa.
- La dimensió social del robot no solament contribueix a disminuir mitjançant la seva presència l'espai i el temps de solitud necessaris per a la interiorització i l'autoreflexió (el que pot produir al seu torn danys psicològics), sinó que pot induir l'usuari a compartir els pensaments més íntims. Un robot pot arribar a convertir-se en el confident d'una persona amb una interacció social reduïda, i uns tercers que tinguin accés a aquestes confidències es poden aprofitar d'eventuals vulnerabilitats.

Per a Ryan Calo, la llei no està ben equipada per a fer front a la dimensió social del robot:

«El consentiment informat tendeix a derrotar les reclamacions sobre privadesa, i és difícil mesurar el dany en causes sobre privadesa.» (Calo, 2014, pàg. 198)

També cal destacar que els robots poden tenir la seva pròpia internet, amb la qual poden adquirir la informació o les habilitats per a fer la tasca encomanada sense mediació humana (encara que els humans sí que poden introduir coneixement en aquesta xarxa). Aquest era l'objectiu del projecte europeu FP7 RoboEarth (European Commission, 2014), que va concloure el 2017. Es tracta bàsicament d'una xarxa d'informàtica en núvol (*cloud computing*, això és, distribuïda en una xarxa de servidors) i un repositori que conté components de programari, mapes per a navegació, coneixement sobre tasques (per exemple, receptes d'acció, estratègies de manipulació) i models per a reconèixer objectes, tot això en un llenguatge intel·ligible per a les màquines.

És important ser conscients que no solament el PC, el mòbil i el robot poden constituir portaveus de la nostra intimitat: amb la progressiva introducció i difusió d'objectes i utensilis quotidians «intel·ligents», i mitjançant el que s'ha anomenat *Internet of Things* (IoT), s'obren altres microportes que poden deixar

El projecte europeu FP7 RoboEarth

Aquest projecte va continuar en els projectes RoboHow, KnowRob i OpenEase.

sortir dades sobre la nostra localització, moviments comercials o financers, gustos i hàbits, relacions, biometria... Pot ser que conjunts d'aquests artefactes estiguin connectats únicament en l'àmbit local (això és, en un àmbit restringit) per a millorar la seva eficiència i seguretat, però també és molt probable que s'estableixin connexions amb internet, per exemple, per permetre actualitzacions.

2.2.6. L'àmbit de la intimitat: relacions afectives

Els humans tendim a desenvolupar un vincle envers determinats objectes personals, sia pels records associats (per exemple, de la persona que ens els va donar), per la funcionalitat que proporcionen (comoditat en la neteja personal, eficàcia a la cuina, mobilitat...) o per certa propensió a antropomorfitzar objectes pel seu aspecte o comportament. Aquest vincle pot arribar a ser molt més profund amb un robot, ja que és fàcil atribuir de manera natural una personalitat única a aquesta màquina pels motius següents:

- Es mou sovint imprevisiblement, de manera que ens podem veure temptats inconscientment a atribuir-li autonomia i intencionalitat pròpia.
- Té un comportament més ric i variat que altres aparells.
- La interacció amb l'usuari humà és de més alt nivell i comprèn instruccions i rèpliques verbals, comunicació gestual, etc.
- Té un aspecte més antropomorf (o amb forma de mascota), justificable sovint des d'un punt de vista funcional.
- La interacció té lloc de manera regular i freqüent, diàriament o fins i tot permanentment.

Aquest tipus de vincles més estrets pot desenvolupar-se amb robots personals l'aspecte dels quals és manifestament el d'una màquina, i fins i tot amb aparells o programari que no tenen la presència física d'un robot però sí que exerceixen un paper destacat en el dia a dia, com els assistents personals que ens mantenen informats sobre el que ens interessa, estan al corrent de la nostra agenda i ens recorden les cites (un simple altaveu que dona veu a una IA incorporada). Però fins i tot pot produir-se d'una manera molt més acusada quan hi ha una dependència física del robot o quan la capacitat de discerniment de l'usuari sobre la veritable naturalesa del robot és condicionada per l'edat o altres factors. Els robots que s'ocupen precisament de la cura de persones, els robots assistencials o cuidadors, susciten un debat ètic més profund a causa de la vulnerabilitat dels usuaris. A Torras (2019) es revisen els èxits de la robòtica assistencial i es presenta el material educatiu emprat en un curs universitari

sobre ètica en robòtica i IA social. En el cas de persones dependents per edat o discapacitat, ¿és desitjable que l'usuari tingui un vincle emocional amb el robot? Podria ser-ho en els casos següents:

- Situacions en què el robot actua com un substitut temporal i terapèutic per practicar formes de comunicació interpersonal saludables sempre sota la supervisió d'un terapeuta. Els avantatges del robot són que pot modelar incansablement diferents comportaments i no jutja ni critica l'usuari, per la qual cosa se sent menys cohibit en la interacció.
- Situacions en què un vincle emocional pot propiciar un comportament saludable del pacient, com prendre puntualment medicaments, animar-se a moure's més i fer més exercici físic, fomentar la inquietud intel·lectual, etc.

No obstant això, el vincle emocional també pot ser contraproductiu pels motius següents:

- La relació es torna obsessiva i/o excloent.
- La pèrdua del robot provoca sensacions de duel similars a la de la pèrdua d'un familiar o ésser estimat.
- Substitueix les relacions afectives entre humans a llarg termini (no es basen en la reciprocitat, sinó que són transaccions).
- Terceres persones (per exemple, els mateixos fabricants del robot) s'aprofiten d'aquest vincle per a manipular l'usuari (per exemple, per a induir-lo a adquirir millores o complements del robot o altres productes de la companyia). Referent a això, Matthias Scheutz adverteix del perill que comporten els llaços emocionals unidireccionals entre humans i robots socials: les dependències psicològiques que creen poden

«tenir serioses conseqüències per a les societats humanes, perquè poden ser explotades a gran escala» (Scheutz, 2014, pàg. 216).

La regulació legal és necessària; tal com afirma Blay Whitby (2014, pàg. 234),

«és poc probable que els problemes socials de l'ètica dels robots sigui resolta deixant que els mercats decideixin lliurement.»

Aquests factors negatius tenen una significació més acusada en el cas de col·lectius vulnerables com infants, ancians, persones amb discapacitat física o mental, o persones amb dificultats d'integració social. És just que el dissenyador i fabricant de robots cuidadors hagi de tenir especialment present la necessitat de preservar la dignitat humana per a aquests grups humans (Whitby, 2014). Però tampoc no cal formar part d'un d'aquests col·lectius per a ser una víctima potencial de la manipulació per part d'un robot de formes que no són

concebibles per a un altre ésser humà. Matthias Scheutz (2014, pàg. 218) apunta que una manera radical d'evitar-ho és equipar-lo, tal com va fer l'evolució amb nosaltres, «amb un sistema emocional que aconsegueix un equilibri entre el benestar individual i un comportament acceptable socialment». Però l'objectiu d'arribar a implementar emocions genuïnes en un sistema artificial no sembla estar a l'abast del desenvolupament científic i tecnològic, almenys no a curt termini. En canvi, Meacham i Studley (2017) no estan tan preocupats per l'engany subjacent en un robot cuidador quant a l'autenticitat de la seva predisposició envers l'usuari: per a ells, el fet rellevant és que l'atenció dispensada pel robot creï un ambient favorable per a la cura o recuperació del pacient. Suggereixen que fins i tot una generació que hagi crescut amb aquesta tecnologia no tendirà necessàriament a confondre robots que tenen una resposta emocional sofisticada (sense ser autèntica) amb éssers vius, però sí que podrà atribuir-los un tercer estat entre la vida i la no-vida.

Conforme els robots vagin aconseguint major sofisticació i antropomorfisme, hi haurà espai per a un altre tipus d'interacció en una esfera més personal i íntima. Cal preguntar-se si és legítim mantenir relacions sexuals amb un robot i si s'han de crear robots amb aquesta finalitat, o si el vincle emocional pot arribar a l'enamorament i si es pot preveure la unió legal en matrimoni entre un ésser humà i un robot. Aquestes qüestions no són matèria de mera i hipotètica elucubració, sinó que ja comencen a ser realitat: la indústria fa una inversió important en nines (i ninots) sexuals cada vegada més avançats, i també s'ha donat el cas d'un japonès, Akihiko Kondo, que el 4 de novembre de 2018 es va casar amb un holograma de la cantant virtual Hatsune Miku d'un dispositiu assistent personal de Gatebox (Jiménez de la Fuente, 2019). El matrimoni no té validesa legal, però la companyia ja ha expedit desenes de certificats matrimonials entre els seus clients i els seus productes.

Un robot pot arribar a convertir-se en un company sexual altament gratificant, ja que pot ser modelat físicament i per caràcter al gust de l'usuari, i és incansable, està sempre disponible i predisposat. Els seus defensors sostenen que poden complir un rol social de gran importància:

- Poden augmentar el nivell de satisfacció sexual del conjunt de la població i, per tant, contribuir a una societat més relaxada, menys frustrada i violenta.
- Poden proporcionar satisfacció sexual a persones amb dificultats físiques i/o socials.
- Poden reduir el tràfic de persones i la seva explotació sexual.

Per contra, els seus detractors creuen que els efectes negatius que poden tenir no són menors:

- Poden provocar la deterioració de les relacions interpersonals, ja que afebleixen la tolerància a opinions negatives o diferents de la nostra i contribueixen a la pèrdua de la riquesa de matisos (evidentment, un robot també pot ser programat per a expressar alguna negativa de tant en tant, però no deixarà de ser un comportament programat).
- Amb relació al punt anterior, poden fomentar una actitud de menyspreu envers el gènere representat pel robot, majoritàriament femení.
- Poden facilitar l'aïllament social (encara que també és factible que fomentin les relacions entre usuaris i propietaris de robots de manera similar a les mascotes amb els seus propietaris).

David Levy (2014), un ferm defensor dels robots sexuals, estableix paral·lelismes entre aquests i altres productes tecnològics que s'han desenvolupat per a l'autosatisfacció sexual, com nines inflables i vibradors, i els considera com la lògica evolució i culminació d'aquests i, per tant, no objectables moralment. Considera que l'estigma social de llogar un robot sexual seria menor que el de ser client de treballadors sexuals humans, i que la percepció del robot com un objecte evitaria sensacions d'infidelitat en una parella humana en què un dels membres (o tots dos) decidís utilitzar-lo. No obstant això, creu que no cal obviar el dilema ètic que es planteja amb els treballadors humans del sexe (els que exerceixen la professió voluntàriament), que es quedarien a l'atur, i amb els mateixos robots si en el futur arriben a adquirir quelcom semblant a la consciència. L'intercanvi sexual és sens dubte un camí cap a la familiaritat i la dependència, però Whitby qüestiona que també ho sigui necessàriament per a l'amor, ja que un requisit per a aquest és la reciprocitat i ja hem vist que poder replicar la capacitat emocional necessària en un robot és, si més no, qüestionable.

Més enllà de la gratificació sexual, els humans poden desenvolupar algun tipus d'afecte envers un robot que els cuidi o que senzillament els «faci companyia». És evident que, per a robots sexuals, un aspecte humà realista està justificat. Aquests robots tindran èxit si són capaços de superar la «vall inquietant» (*uncanny valley*) que es produeix en acceptar-los conforme augmenta la seva similitud amb un ésser humà (hi ha andròides molt semblants que generen inquietud i rebuig). Però el dissenyador (i el fabricant) poden apostar per l'honestedat i crear robots que no dissimulen que són màquines pel seu aspecte i que resulten així i tot molt atractius, per a altres finalitats com l'assistència, l'educació o la companyia. En la seva novel·la *La mutació sentimental*, la professora Carme Torras (2018, pàg. 246) fonamenta l'atracció que la protagonista Celia sent pel seu robot personal en el

«seu comportament lleial, digne de confiança i previsible, reforçat per la seva aparença no enganyosa de màquina»⁷.

Reflexió

L'autor apunta al fet que avui dia cadenes de farmàcies respectables tenen vibradors a les prestatgeries (vegeu també David Levy, *Love and sex with robots: the evolution of human-robot relationships*, Harper Perennial, 2008).

⁽⁷⁾ La citació és de l'apèndix dedicat a la discussió de temes i qüestions per a grups de lectura, solament en la versió anglesa.

Aquestes característiques no estan renyides amb el fet que el robot pugui proporcionar una conversa estimulante o una interacció enriquidora. Elder (2017) estableix el paral·lisme amb la moneda falsificada (que intenta passar per autèntica) en els «dinners» del Monopoly, el valor de la qual està restringida clarament al joc i no pretén enganyar.

Una altra qüestió que es planteja és la del llegat: el robot personal que «sobreviu» a l'amo, que n'atresora els records mantenint, potser de manera anàloga a un diari íntim, una part del que era aquesta persona. Com disposarà aquesta persona d'aquest llegat encara en vida? Quines opcions legals tindrà? Podrà minvar la seva herència a favor del manteniment del robot? Fins a quin punt hi podran interferir els hereus? Tindran algun dret sobre el llegat? Qüestions similars ja es plantegen avui dia respecte a la presència digital de les persones a internet.

3. Iniciatives legals

La resolució d'un conflicte generat per un robot o per una IA incorporada requerirà moltes vegades la intervenció dels tribunals. En aquesta tercera part es descriurà com les eines de l'IA ja formen part del món jurídic i pot ser que algun dia ajudin a resoldre aquest tipus de casos. La problemàtica més immediata està relacionada amb la seguretat, i en aquest sentit parlarem dels fonaments de la responsabilitat legal relacionada amb els robots. També es descriuran els primers avenços per a establir les lleis pertinents tant a Europa com en altres països amb prou desenvolupament tecnològic, i el que ja s'ha avançat en el terreny de la normativa, tant en la robòtica industrial com en la de serveis. Finalment, ens permetrem proposar un breu exercici especulatiu sobre possibles drets futurs dels mateixos robots.

3.1. La intel·ligència artificial i el món de les lleis

Els robots i l'IA ja han entrat al món jurídic: els primers, com es veurà en el subapartat 3.2.1, en la mesura que són elements involucrats en accidents mortals. Estem parlant de robots industrials i robots experimentals que han entrat en contacte amb un grup de persones reduït. Conforme s'expandeixin en nombre i es diversifiquin en atribucions, es multiplicaran els casos legals, no necessàriament per danys físics sinó per altres tipus de danys (a la propietat, a la privadesa, etc.). En els subapartats següents veurem com comencen a fer notar la seva presència en l'àmbit jurídic sobretot de manera preventiva, això és, com el món de les lleis comença a preparar-se pel que serà sens dubte un dels majors canvis tecnosociològics experimentats per les civilitzacions humanes. L'IA incorporada té una presència positiva al món jurídic al marge de la seva possible utilització delictiva en accions d'espionatge, d'atac a la propietat o *hacking* de sistemes el control dels quals depèn de programari, això és, d'accions punibles que comencen a aparèixer a les sales de judici: aquest subapartat està dedicat precisament a repassar quines eines i tècniques d'IA poden ser i són d'utilitat en el món de les lleis.

El programari d'assistència a lletrats adopta formes cada vegada més sofisticades i útils gràcies a l'IA. Cal destacar el canvi de paradigma que s'ha produït a les últimes dècades. A l'inici van ser els sistemes experts, amb les seves representacions del coneixement i inferència basats en regles, que van tenir una gran difusió dels anys setanta als noranta del segle passat. Això responia a l'objectiu de «desenvolupar models formals d'argumentació legal en una forma processable per ordinador i modelitzar computacionalment la legislació i les regles legals» (Surden, 2019). L'article de Buchanan i Headrick (1970) sol ser referenciat com el primer treball seriós que explora les possibilitats de l'IA en el món jurídic, i apunta les línies mestres del que serien els seus objectius i mètodes durant l'esmentat període: des de la cerca de paraules clau en textos

legals, augmentada amb la cerca associativa (coocurrència de paraules en el mateix text), fins als sistemes de raonament legal basats en regles. L'article és més aviat una compilació de *desiderata* i apunta les dificultats per a formalitzar el raonament jurídic, encara que subratlla l'interès que té l'IA per abocar precisament alguna llum sobre com es produeix aquest raonament. No obstant això, a partir del canvi de mil·lenni hi ha hagut un canvi d'orientació cap als processos basats en aprenentatge automàtic seguint les tendències de l'IA en general. A Surden (2019) s'apunta a una classificació de l'ús de l'IA en el món jurídic atenent a tres tipus d'usuaris: els administradors de la llei (legisladors, jutges, oficials administratius i policies), els professionals de la justícia (advocats bàsicament) i els afectats o governats per la llei (ciutadans, negocis, empreses i corporacions). Per exemple, els jutges poden sotmetre a un sistema d'IA l'avaluació del risc de reincidència d'un processat per decretar-ne o no la llibertat condicional. També la policia pot utilitzar tècniques d'aprenentatge automàtic per a determinar les zones més conflictives de la ciutat i incrementar la seva presència en aquestes zones. Una altra tecnologia d'ús cada vegada més estès en l'àmbit policial és el reconeixement facial a partir d'imatges recollides per càmeres. Pel que fa a usuaris circumstancials de la llei, hi ha els sistemes experts que utilitzen algunes companyies per a comprovar si determinades activitats, per exemple d'importació-exportació, s'ajusten a la llei, o els simples sistemes experts (sovint en forma de *chatbots* o programes conversacionals) que una persona sense formació legal pot utilitzar per a resoldre els seus dubtes sobre qüestions legals bàsiques (Surden, 2019, pàg. 1335). Però el que té més interès en aquest curs són les eines que l'IA ofereix al segon dels col·lectius esmentats. Potser el més destacat són els sistemes que revisen milers de documents per detectar els que són rellevants per al cas en curs (abans del judici): les anomenades tècniques de codificació predictiva, que usen aprenentatge automàtic i representació del coneixement per a automatitzar el procés de determinar la rellevància jurídica d'un document (per a un cas determinat). Un exemple que s'esmenta a Frey i Osborne és el

«sistema Clearwell de Symantec, que usa anàlisi de llenguatge per a identificar conceptes generals en documents, que pot presentar els resultats gràficament i que ha demostrat ser capaç d'analitzar i ordenar més de 570.000 documents en dos dies»⁸. (Frey i Osborne, 2013, pàg.17)

⁽⁸⁾ Els autors citen al seu torn Markoff (2011).

L'IA és realment eficient per a dur a terme aquestes tasques tedioses, que implicarien una quantitat insostenible d'hores-persona i poden servir sobretot per a eliminar informació irrellevant, però la decisió final sobre la rellevància d'un document continua essent d'un advocat humà, ja que involucra la *comprensió* de la llei i els fets relacionats amb el cas, i implica raons estratègiques i abstraccions que encara estan fora de l'abast de l'IA. La predicció de les probabilitats associades als diferents possibles resultats d'un judici és una altra aplicació en què les tècniques d'IA poden ajudar. Això pot servir a l'advocat per a aconsellar el seu client sobre les seves possibilitats i si interessa arribar a un acord amb l'altra part.

A dia d'avui els sistemes experts utilitzats en l'àmbit legal disposen d'una de les quatre arquitectures següents: sistemes basats en regles amb raonament deductiu, models d'anàlisi basada en casos (això és, per analogia amb casos similars), xarxes neuronals i lògica difusa (Wikipedia, «Legal expert system», 2019; Wikipedia, «Legal informatics», 2019, epígraf «2.2. Artificial intelligence»). Els reptes a què s'enfronten són la complexitat inherent al dret, les limitacions de la resolució de problemes per ordinador, la dificultat de representar el coneixement legal (per les seves imprecisions, pel fet que no es produeixen sempre els mateixos resultats –d'un judici– en condicions similars, etc.), les restriccions autoimposades en temes jurídics i de jurisdicció (això és, es creen sistemes experts d'aplicabilitat limitada), i possibles interpretacions errònies dels resultats del sistema expert en ser interpretats per un ignorant en la matèria (Wikipedia, «Legal expert system», 2019). També cal destacar que la seva aplicació a un camp ampli però concret com les lleis ha reportat alguns resultats i desenvolupaments interessants a l'IA més genèrica.

3.2. Responsabilitat civil i penal dels actors implicats en el desenvolupament del robot intel·ligent

3.2.1. Accidents: els robots poden causar dany

Desgraciadament, la història de la robòtica, encara breu, ja disposa d'una pàgina negra d'accidents amb víctimes mortals causades per robots. Ja el 1979 es va produir el que es considera primer accident mortal ocasionat per un robot (Wikipedia, «Robot fatality», 2019), quan Robert Williams, un treballador de la planta de modelatge de la Ford Motor Company, a Míchigan, va ser colpejat al cap pel braç mecànic d'un robot d'una tona de pes mentre retirava manualment una peça a causa d'un funcionament aparentment defectuós del robot. El 1983 l'empresa fabricant del robot, Litton Industries, va ser condemnada per un jurat popular a pagar, mitjançant la seva divisió Unit Handling Systems, 10 milions de dòlars als hereus de Williams per negligència «en el disseny, fabricació i subministrament del sistema d'emmagatzematge, i per desistiment de l'avís [als operadors del sistema] dels riscos del treball a l'àrea d'emmagatzematge». L'empresa va arribar a un acord amb els familiars de Williams per un import no revelat a canvi de no admetre negligència. Posteriorment, Litton va intentar recuperar les costes del judici i una indemnització, i va demandar Ford per no haver enviat Williams a l'entrenament ofert per Litton i per permetre l'entrada del treballador a un àrea d'emmagatzematge sense activar el sistema de parada. Però el tribunal va desestimar la demanda perquè ja havia arribat a un acord amb els hereus de Williams.

El 1981 es va produir el primer accident mortal en una fàbrica de robots, quan Kenji Urada, un treballador de la Kawasaki Heavy Industries de la planta de la ciutat japonesa d'Akashi, va saltar la tanca de seguretat per comprovar el funcionament erroni d'un robot (Wikipedia, «Kenji Urada», 2019). Va activar distretament el robot, que casualment va aixafar el treballador contra una mà-

quina rectificadora i li causà la mort. Els altres treballadors, que no estaven familiaritzats amb l'operació del robot, van ser incapaços de desconnectar-lo. La recerca subsegüent va posar en relleu aquesta falta de formació i la falta de regulació d'aquestes màquines. El robot involucrat en l'accident va ser retirat, i es van construir tanques de seguretat de l'alçada d'una persona al voltant dels altres robots de la planta.

Al seu moment aquests dos casos van posar en relleu una sèrie de qüestions fonamentals sobre els robots industrials:

1) El robot industrial és una màquina potencialment perillosa, que pot causar greus ferides i fins i tot la mort, a causa de la seva envergadura i les trajectòries complexes que pot seguir el braç mecànic. Una seqüència mal programada pot originar un moviment imprevist per l'operador, amb el consegüent risc de col·lisió.

2) La normativa que introduïa mesures de seguretat com la programació a velocitat reduïda i el polsador d'home mort, la tanca de seguretat amb detecció d'intrusos i la parada automàtica, el polsador de parada d'emergència en un lloc visible i accessible, la formació adequada d'operaris, etc. es va regular *a posteriori*, una vegada ja s'havien produït els accidents.

3) La responsabilitat legal pels accidents pot recaure tant en el fabricant del robot com en l'empresa usuària. La legislació ha de delimitar clarament les atribucions de cada actor involucrat en matèria de seguretat, i la recerca judicial ha de dilucidar el grau de compliment de la normativa per les parts, inclosa l'eventual víctima.

També s'han produït accidents fatals fora de l'àmbit industrial. El 2007 un canó robòtic antiaeri (semiautònom) de l'exèrcit sud-africà va matar per accident nou soldats participants de l'exercici de tir i en va ferir catorze (Shachtman, 2007), i el 2018 un vehicle Tesla en mode autopilot va col·lidir contra la mitgera i va causar la mort del conductor (Brown, 2019), que aparentment no havia reaccionat malgrat els avisos del sistema (durant els 6 segons previs a l'accident no tenia les mans posades al volant).

Fa dècades que els robots industrials funcionen. Cal preguntar-se si l'experiència acumulada en matèria de seguretat és extrapolable a altres àmbits fora de les fàbriques. Ho és únicament de manera molt limitada: generalment, el robot industrial s'ha concebut fins al moment com una màquina que fa tasques perilloses, avorrides o en ambients tòxics, i que opera, per tant, de manera autònoma, sense contacte amb humans. Totes les mesures de seguretat van encaminades precisament a evitar aquest contacte mentre el robot està en funcionament. No es pot procedir de la mateixa manera amb robots la missió dels quals comporta precisament la proximitat o fins i tot el contacte humà. De fet, en la mateixa indústria està cobrant força el concepte de *cobot*, robot

que treballa de manera col·laborativa amb l'operari humà (vegeu el subapartat 2.1). Però dels tres punts anteriors sí que es poden extreure algunes conclusions vàlides per a altres tipus de robots:

1) A diferència d'altres màquines que poden causar greus accidents, com premses o cisalladores, on el recorregut dels elements mòbils és sempre el mateix, els braços robots poden accedir a qualsevol punt del seu espai de treball fent trajectòries que poden seguir una seqüència programada més o menys complexa o que varien en resposta a estímuls externs. En el cas dels robots mòbils, l'accés és pràcticament il·limitat. Com que no poden ser confinats a causa de les tasques que se suposa que han de fer, les mesures de seguretat passen pel disseny dels mateixos robots, tant de la seva constitució física com de la seva programació i mecanismes de control. El disseny contribueix a la seguretat intrínseca del robot, que és solament un aspecte (encara que sens dubte el més important) de l'ètica *implícita* d'aquestes màquines (recordeu el subapartat 2.1).

2) Els accidents es produeixen, però una normativa desenvolupada amb prou previsió pot reduir-ne la freqüència i pal·liar-ne els efectes. El dany físic i/o econòmic no és l'única conseqüència negativa que pot ocasionar un robot en la seva interacció amb humans, ni els accidents l'única manera en què aquests efectes perjudicials es poden produir. Anticipar-se a les possibles contingències requereix un esforç d'imaginació i de sentit comú (dues habilitats cognitives en què els humans encara superem les màquines intel·ligents), i un coneixement precís i actualitzat de les possibilitats reals dels robots i de la naturalesa humana. Les lleis i recomanacions que es derivin d'aquest esforç no seran l'última paraula, sinó que hauran de ser revisades i actualitzades d'acord amb la pràctica real i de les eventualitats que s'observin i amb els avenços en mobilitat, destresa i habilitats cognitives que es vagin desenvolupant en els robots.

3) De moment, i encara durant algun temps, la responsabilitat legal per una eventualitat perniciosa generada per un robot recaurà en algun dels nombrosos actors humans involucrats en la interacció amb el robot, però no en la mateixa màquina, ja que no està dotada dels requisits cognitius necessaris per a fer-se responsable de les seves accions. De nou, pel que fa als robots en general, això no concerneix solament contingències com accidents, sinó que es fa extensiu a un altre tipus de conseqüències adverses, com danys psicològics o morals. Algunes veus es deixen sentir en el sentit que és possible atribuir *personalitat legal* a màquines sense que hagin arribat a adquirir atributs com una consciència similar a la humana, ja que, de fet, aquesta figura existeix per a entitats com corporacions o animals no racionals.

3.2.2. Un marc legal per a robots

Prescindint d'entrada de la naturalesa moral d'un robot (això és, si es pot considerar o no un agent moral), la veritat és que tant les accions que fa com les que rep poden tenir una dimensió moral pels possibles danys i perjudicis ocasionats per aquestes accions sobre persones o propietats (inclosa la seva inte-

gritat). Les lleis hauran de complir la comesa de protegir els drets de les persones que siguin afectades per les accions d'un robot i, atesos els mecanismes de què disposen, ho faran de manera molt més efectiva que les recomanacions, manifestos, declaracions o altres procediments que puguin reflectir voluntats normatives o reguladores de la presència i difusió de la robòtica i l'IA en la societat. Caldria preguntar-se si el marc legal actual ja és suficient, de manera que la jurisprudència que es generi bastarà per a afinar l'especificitat associada a aquestes tecnologies basant-se en el bon criteri i sentit comú dels jutges que es vagin enfrontant als casos que es produeixin. La resposta és afirmativa solament per a un reduït nombre de casos. Per tant, el febrer de 2017 el Parlament Europeu va emetre una resolució sobre regles de dret civil en robòtica (que comentarem en el subapartat 3.3) arran de les conclusions del projecte europeu RoboLaw (2014) (vegeu el subapartat 3.3.1). Aquesta resolució preveu les particularitats del fenomen robòtic i de l'IA, i serveix de marc de referència comuna per a les iniciatives legislatives dels estats membres. Abans de procedir a comentar els aspectes més destacats d'aquesta resolució i d'altres iniciatives legals en l'àmbit extracomunitari, és just repassar alguns conceptes teòrics, bàsicament sobre responsabilitat civil i penal, i com s'apliquen al cas del robot en diferents fases de desenvolupament cognitiu. Hem seguit majoritàriament el text aclaridor de Peter Asaro (2014), que, encara que parteix de la tradició jurídica nord-americana, té un enunciat de conceptes, principis i raonaments de caràcter més general.

Robots no autònoms o semiautònoms

En l'ús més simple, en què el robot es limita a executar un programa preestablert de fàbrica, els casos que es presentin quedaran coberts per la legislació civil i penal de la responsabilitat pel producte⁹ (Asaro, 2014). El robot és tractat com qualsevol altre producte manufacturat tant pel que fa als danys causats per les seves accions com als danys (inclosa la substracció) que tercers infligeixin sobre la màquina (el que es consideraria com a danys sobre la propietat). Concentrant-nos en el primer tipus, el dany causat per un robot pot deure's a una concepció i disseny descarat, una fabricació defectuosa, una programació errònia o un ús no adequat. En tota aquesta cadena, que involucra científics, tecnòlegs o dissenyadors, fabricants, programadors, distribuïdors i usuaris, no s'ha suposat una intencionalitat criminal; altrament, caldria incloure-hi també sabotejadors i *hackers*, i estaríem parlant de responsabilitat penal pels danys causats. Circumscriuint-nos a la responsabilitat civil associada a possibles danys causats de manera no intencionada, l'objectiu és determinar qui ha de compensar la víctima (els tribunals han de decidir, a més, la quantia d'aquesta compensació). Per simplificar, considerarem únicament dos actors, el fabricant i l'usuari, el primer dels quals concep i fabrica el robot, i el segon dels quals l'utilitza (pel que fa a la programació, cal distingir entre la programació de base que ve de fàbrica i la configuració particular de l'usuari, i les

⁽⁹⁾ O responsabilitat sobre el producte o responsabilitat per defectes del producte.

modificacions que aquest faci sobre la programació original). El fabricant ha de respondre per dos possibles tipus d'errors associats a negligències en el disseny i la fabricació:

- Error d'avisar (l'usuari sobre riscos previsibles associats a l'ús del producte).
- Error de prestar l'atenció i dedicació suficients a determinar els potencials riscos del producte.

Determinar tots els riscos associats a un producte no és fàcil, sobretot quan es tracta d'un producte nou, complex i en ràpida evolució com el robot. Els fabricants sovint s'escuden en el que es considera la pràctica *normal* del sector (sigui una normalització explícita com les normes ANSI o ISO, o la pràctica implícita), o ho prenen com a referència. En el cas dels robots no hi ha encara prou experiència per a consolidar el que es podria considerar com a pràctica habitual, i aquesta podria ser precisament una de les comeses més útils de la roboètica: imaginar els possibles riscos i desenvolupar mesures de precaució. Si la negligència del fabricant és fruit d'una manifesta actitud d'eludir qualsevol estudi dels potencials riscos del seu producte, pot ser qualificada de criminal: és un cas de *mens rea*, on

«l'estat de culpabilitat de la ment és d'ignorància, sigui voluntària o per poc raonable» (Asaro, 2014, pàg. 172).

Estadis encara més avançats de culpabilitat són els de temeritat o imprudència (previsió de possibles riscos o danys, encara que sense la intenció que alguna víctima potencial sigui danyada) i els de consciència de l'existència d'un risc real (de nou sense la intenció de causar dany). Malgrat tot, aquests casos se solen resoldre per la via civil. Cal assenyalar també que la responsabilitat causal no sempre coincideix amb la legal, i les legislacions s'esforcen perquè es compleixi l'objectiu de compensació de la víctima pel dany causat, sia amb la repartició de responsabilitats o per atribuir responsabilitat estricta a l'actor que està en condicions més favorables de reparar la víctima. En el cas del robot, on conflueixen diversos components (possiblement de diferents fabricants), la casuística de possibles interaccions entre els components pot ser molt complexa. La construcció d'altres sistemes de complexitat elevada, com les aeronaus, pot servir d'inspiració per al cas del robot.

La responsabilitat del dissenyador, enginyer o fabricant acaba on comença la de l'usuari; generalment, és determinada per un ús indegut del producte. De fet, amb determinats productes perillosos, com vehicles de motor o armes, la responsabilitat recau totalment en l'usuari, i la del fabricant queda restringida a possibles defectes de fabricació. Un robot podria entrar en aquesta categoria depenent de les seves característiques. No és realista que un fabricant pugui anticipar tots els possibles usos del seu producte, i com a regla general pot valer que com menys previsible sigui un cert ús del producte menor serà la responsabilitat del fabricant (i major la de l'usuari). En el cas del robot, es poden interpretar com a ús tant les tasques a què es destina (que depenen al seu torn

de com s'hagi programat el robot) com el context o entorn en què es desenvolupa la seva activitat, que pot ser totalment inadequat. Un usuari pot ignorar la resposta del seu robot en una situació determinada, la qual cosa pot ser en certa manera un atenuant si es produeix algun dany. Però si posteriorment es produeix un dany en una situació similar, l'usuari estarà en una posició de *coneixement dolós* (White i Baum, 2017). Serà molt difícil evitar que usuaris avesats en programació o manetes del bricolatge tunegin els seus robots, encara que és factible sotmetre'ls a una revisió periòdica, una mena d'ITR (inspecció tècnica de robots) en què es comprovi que les modificacions s'ajusten a la legalitat. Atribuir una responsabilitat legal estricta al fabricant, d'altra banda, pot comportar una notable trava a la innovació i desenvolupament, i a la vegada limitar la predisposició d'aquest fabricant a fer que el seu programari sigui de codi obert, la qual cosa posa al seu torn obstacles al desenvolupament d'aplicacions per part de la comunitat d'usuaris, que podrien redundar en benefici de tothom. D'altra banda, conforme avanci la capacitat dels robots de processar el llenguatge natural, la interacció entre robots i usuaris serà més verbal (oral), la qual cosa obrirà les portes a tots els equívocs i imprecisions del llenguatge humà, i les possibles interpretacions errònies de les ordres. En alguns casos serà complex dilucidar si el comportament erroni del robot es deu a un error en el programari d'interpretació del llenguatge o al fet que l'usuari no s'ha expressat amb prou claredat.

Robots autònoms

El desenvolupament de les tècniques d'IA permetrà dotar els robots de certa capacitat de raonament en algun moment. Seran capaços de determinar un curs d'acció i de triar una opció entre diverses per complir formulats a més alt nivell, com «mantingues la casa neta» o «cuida de l'avi Pancratius». El grau de sofisticació del mecanisme de presa de decisions pot arribar a un extrem en què sigui molt complex, per no dir impossible, atribuir la responsabilitat causal d'una determinada acció amb resultats nocius al fabricant o a l'usuari, i ens vegem temptats a fer responsable el mateix robot dels seus actes. No obstant això, no s'ha d'oblidar que en última instància és un ésser humà (o un grup d'éssers humans) qui va decidir desenvolupar aquest robot, i que també és un ésser humà qui va decidir adquirir-lo i usar-lo. El robot en qüestió no és un ésser independent, sinó que té un propietari. És possible que aquest robot hagi pres una decisió de manera autònoma, en tant que se li pot reconèixer capacitat de raonament, i que aquesta decisió no sigui fruit d'una falta negligent de previsió per part del fabricant ni hi hagi intencionalitat dolosa per part de l'usuari al moment d'ordenar-li les tasques. No obstant això, fins i tot admetent una responsabilitat causal en el robot, no s'haurà de considerar necessàriament que incorre en una responsabilitat moral o legal. En el context dels robots militars, Lokhorst i Van den Hoven (2014, pàg. 151) analitzen la responsabilitat moral d'un robot que compleix ordres d'un superior en els dos casos de robot no deliberatiu i deliberatiu. Per a això, apliquen el formalisme lògic de l'STIT (*seeing to it that*, «procurar que»: un agent B procura que [es produeixi, passi, es compleixi] P [una proposició corresponent a aquest efecte

de certa acció]) i la variant del DSTIT (*deliberatively seeing to it that*, on «deliberativament» comporta que també es podria optar per «no P», és a dir, que hi ha capacitat d'escollir). En el cas d'un robot no deliberatiu (és a dir, en el cas exposat en el subapartat 3.2.2.1), s'aplicarien les tradicionals màximes legals de *qui facit per alium facit per se* (això és, que actuar mitjançant uns altres, subordinats d'alguna manera, equival a actuar un mateix) i *respondeat superior* (l'amo [ha de] respon[dre]). En el cas de robots deliberatius, els autors consideren els casos d'«elecció forçada» i, per tant, de «responsabilitat moral forçada», i estableixen paral·lelismes amb soldats que, si bé no són responsables moralment de qüestionar la legalitat de la mobilització, sí que poden ser-ho sobre la decisió de complir o no ordres específiques que per exemple contravinguin el DIH. El conflicte d'aquesta responsabilitat amb l'obediència deguda va ser objecte dels famosos judicis de Nuremberg, on quedà de manifest que la segona no era eximent davant la primera. No obstant això, precisament en el context bèl·lic els autors consideren qüestionable que s'arribin a incorporar robots que en algun moment puguin decantar-se per la insubordinació.

En un context menys específic, Asaro (2014) també considera la responsabilitat civil subsidiària dels propietaris de robots establint paral·lelismes entre els propietaris d'animals domèstics i els d'animals salvatges, així com entre ocupador i treballador, oficial i soldat, i fins i tot mestre i esclau. En el cas de l'animal domèstic, que l'autor considera comparable amb el d'un robot amb capacitats deliberatives però amb certa experiència quant a la seva utilització, amb un cert rodatge en societat, el seu comportament és bastant previsible en principi. Si en algun moment causa un dany (actuant autònomament), el propietari de l'animal es considerarà responsable civil, i si l'animal actua agressivament perquè el propietari ha tingut una actitud criminalment negligent, temerària o fins i tot dolosa, aquest serà responsable penalment del dany causat. Les intencions de l'animal no es consideren rellevants i, en canvi, sí que s'hi consideren les del propietari, encara que també és cert que un comportament inusualment erràtic o inesperat de l'animal tendeix a disminuir la *mens rea* del propietari (si bé no l'eximeix de la seva responsabilitat). En cas de tinença d'animals salvatges, si poden comportar un cert perill per a les persones, la responsabilitat legal del propietari és major, i si aquest no és capaç de mantenir-los en un recinte controlat (és a dir, si se li escapen) pot incórrer automàticament en una temeritat criminal. El paral·lelisme en el cas de la robòtica seria amb robots que per les seves característiques físiques i per la seva índole experimental o altament modificada poden comportar un perill potencial per al veïnat. L'autor posa l'exemple del propietari que decideix dotar aquest robot d'un arma: encara que el robot no arribi a ferir ningú ni a destruir cap propietat, es considerarà que el propietari incorre en temeritat criminal de la mateixa manera que un conductor en estat d'embriaguesa que es posa al volant d'un vehicle encara que no causi cap accident.

La propietat o l'ús de determinats robots es poden sotmetre a un sistema de llicències de manera anàloga al permís de conduir que és necessari aconseguir per a conduir un cotxe o a una llicència d'armes o de maneig d'explosius (amb els requeriments consegüents d'edat, estabilitat psicològica i coneixements tècnics).

El grau d'autonomia d'un robot podria arribar a l'extrem que tingués objectius propis. En aquest cas es podrien considerar paral·lelismes amb el binomi empresari-empleat. Asaro esmenta un exemple aclaridor sobre l'empleat d'una empresa de transports: si el conductor es desvia de la ruta prevista i causa un accident, la responsabilitat recau en l'empresari si el desviament ha estat forçat (per unes obres, per exemple) i, en canvi, recau en el conductor si ha estat per a atendre assumptes personals (per a visitar un amic de camí, posem per cas). Aquesta dicotomia *detour* enfront de *frolicking* es podria aplicar també al robot pel que es refereix a l'atribució de responsabilitats en el cas hipotètic en què els robots arribin a tenir assumptes personals per atendre.

Podrien els robots tenir responsabilitat penal àdhuc sense ser agents morals plens? A diferència de la responsabilitat civil, en la responsabilitat penal la intenció sí que és rellevant. De nou Asaro estableix paral·lelismes que poden abocar alguna llum sobre com tractar jurídicament un cert tipus de robots avançats, aquesta vegada amb infants i persones amb trastorns psíquics, que es poden considerar legalment quasipersones pel que fa a les seves responsabilitats (encara que siguin persones de ple dret pel que es refereix a eventuais crims sobre ells; de fet, gaudeixen d'una protecció especial en aquest sentit). Encara més, fins i tot un adult estable psicològicament pot al·legar un trastorn mental transitori per beneficiar-se d'atenuants o eximents en un acte punible que hagi comès. En aquest sentit un robot dotat de moralitat funcional (vegeu el subapartat 2.1) es podrà considerar legalment una quasipersona en la mesura que no hagi adquirit personalitat jurídica plena.

Recordem també que la responsabilitat penal comporta el càstig de l'infractor a diferència de la responsabilitat civil, amb la qual es pretén compensar (en general econòmicament) la víctima. La finalitat del càstig és triple: pagar un deute amb la societat (retribució), modificar el comportament de l'infractor perquè no torni a cometre un fet punible (reforma) i buscar l'exemplaritat perquè uns altres infractors potencials s'abstinguin de cometre l'acte punible (efecte dissuasori). Asaro qüestiona que es pugui castigar un robot precisament perquè, mentre no és un agent moral ple, cap d'aquests tres factors no és significatiu realment per a la màquina. Hi ha altres entitats, les corporacions, que sí que gaudeixen de personalitat legal i que també poden ser castigades tant amb la multa, que incideix precisament sobre la seva raó de ser (la maximització del benefici econòmic) com amb la persecució legal dels individus que les integren. Les corporacions són incorpòries (malgrat la paradoxa etimològica), però són capaces de *patir* un càstig. D'altra banda, els robots són incaços de patir àdhuc disposant d'un cos, tret que en algun moment el desenvolupament d'una consciència artificial comporti també la capacitat de patir

Frolicking

Detour es tradueix com a 'desviament', mentre que *frolicking* té connotacions lúdiques, d'entreteniment.

(recordem també que els animals no humans superiors poden patir per la seva capacitat d'experimentar dolor fins i tot no essent conscients en el sentit humà). Però és legítim crear aquesta capacitat de patiment? Lockhorst i Van den Hoven (2014, pàg. 149) no ho creuen així, ja que per a ells el càstig no és una finalitat per si mateix. Entre els tres motius associats al càstig, el de la reforma és sens dubte el més important, i si aquesta fi pot ser aconseguida per altres mitjans (la qual cosa és certament més factible en el cas dels robots, per exemple ajustant-ne la programació) seran sens dubte preferibles al càstig. El patiment també es pot interpretar d'una manera més funcional sense entrar en la seva experiència subjectiva. A Torras (2016b) es discuteix aquesta funcionalitat en el context de l'aprenentatge per reforç (modalitat de programació que es descriu en el subapartat 1.4).

3.3. Iniciatives legals en l'àmbit europeu

3.3.1. Precedents

Les visions ètiques i els costums que constitueixen la moral acceptada en una societat solen precedir la seva formulació precisa i redacció en forma de lleis. En el cas de la robòtica i l'IA passa el mateix amb l'excepció que molts desenvolupaments associats a aquesta tecnologia encara estan més en el terreny d'allò que és factible o previsible que en el d'allò que és real. No hi ha encara una experiència consolidada en l'ús dels robots que pugui inspirar el legislador, però la transformació social que aquesta tecnologia comporta es preveu de tal impacte que molts experts recomanen anticipar-se als fets. Aquí s'imposa el sentit comú, els escenaris que anticipa una certa ciència-ficció de qualitat i les analogies amb altres tecnologies transformadores com l'ordinador personal, internet o la telefonia mòbil (que, de fet, també estan en plena evolució).

L'1 de març de 2012 es va començar a realitzar el projecte del Setè Programa Marc de la Unió Europea («Ciència en societat»), titulat RoboLaw, acrònim de *Regulating emerging robotic technologies in Europe: robotics facing law and ethics*. El seu objectiu principal era elaborar un document sobre les «línies mestres en la regulació de la robòtica» «que contingués suggeriments reguladors per a la Comissió Europea». Els investigadors del projecte van optar per plantejar un estudi per casos tant per la impossibilitat de dissociar la tecnologia robòtica utilitzada d'una aplicació específica com per la impossibilitat de cenyir-se a una tecnologia concreta, ja que en una aplicació n'hi concorren diverses. Els casos estudiats (pel seu interès segons una enquesta popular) comprenien els vehicles autònoms, els robots cirurgians, les pròtesis robòtiques i els robots assistencials. El projecte va traspassar l'àmbit europeu, ja que al mateix any de l'inici van crear una iniciativa conjunta sobre robots, lleis i política amb l'Institut del Dret a Internet de la Universitat de Pequín (Xina), a la qual es va sumar l'Institut de Robòtica Humanoide de la Universitat de Waseda (Japó) (Robolaw Asia, 2012). El projecte va concloure el maig de 2014 i les seves conclusions van ser presentades el 24 de setembre al Parlament Europeu en el

RoboLaw

<http://www.robolaw.eu/index.htm>

taller «Upcoming issues of EU law» (organitzat pel Comitè JURI d'Assumptes Jurídics. L'article presentat es titulà «Regulació de la robòtica: un desafiament per a Europa» (RoboLaw, 2014).

3.3.2. Iniciativa legislativa del Parlament Europeu

La Resolució del Parlament Europeu del 16 de febrer de 2017, amb referència P8_TA(2017)0051, es titula «Regles de dret civil sobre robòtica» (Civil law rules on robotics) i pretén precisament establir les bases d'una futura legislació. Emanava d'una sèrie de principis ètics, entre els quals cal destacar els següents (de manera resumida):

- Atendre els riscos que aquesta tecnologia planteja sobre la seguretat humana, la salut, la llibertat, la privadesa, la integritat, la no-discriminació per raó de gènere, ètnia, creences religioses, ideologia política, etc.
- Actualitzar i complementar el marc legal de la UE amb principis ètics d'acord amb la complexitat dels robots intel·ligents.
- Annexar una carta amb el codi de conducta per a enginyers de robots i comitès d'ètica, i models de llicència per a dissenyadors i usuaris (més endavant es donen algunes indicacions sobre què contemplem aquests codis i llicències).
- Garantir la transparència de les decisions i accions del robot, i suggerir recórrer a mecanismes com la caixa negra dels avions que registrin lectures dels sensors, processos interns i accions executades.
- Proposar la creació d'una Agència Europea per a la Robòtica i l'IA.

Altres punts d'interès inclouen les consideracions següents:

- Conforme creixi l'autonomia dels robots, es podran considerar cada vegada menys simples eines.
- No obstant això, a dia d'avui la responsabilitat pels actes del robot es pot traçar fins a un dels actors humans.
- El marc legal actual és insuficient a causa d'una certa impredictibilitat del comportament del robot.
- Es proposa la creació d'un sistema de registre a escala de la UE.
- Cal posar el focus a complementar i no a substituir els humans, a mantenir el control i vigilar la creació de vincles emocionals.
- S'ha d'afavorir per disseny la recerca, la interconnectivitat requerida, la interoperabilitat, la seguretat i la privadesa.
- Cal una llibertat de flux de dades de manera que es preservi la protecció de dades personals.
- Cal una estandardització i seguretat del tipus ISO/TC 299 Robotics.
- Les proves que es facin en escenaris reals han de seguir el principi de precaució.

Les recomanacions dirigides tant a professionals (enginyers, dissenyadors, comitès d'ètica) com al públic general (consumidors i usuaris de robots) tenen un interès especial, ja que poden inspirar les bases de futures legislacions específiques i poden tenir d'entrada la força normativa d'un codi deontològic. Cal remarcar el caràcter pràctic d'algunes de les recomanacions:

1) Codi de conducta ètica per a enginyers robòtics

- Cal considerar els drets fonamentals.
- Observar el codi de conducta és un acte voluntari.
- La seva elaboració requereix un esforç multidisciplinari.
- Cal considerar les implicacions futures de la tecnologia (no limitar-se als efectes immediats).
- Les entitats de finançament de projectes han de demanar els corresponents informes de seguretat.
- Els humans, no els mateixos robots, són els responsables dels actes dels robots.
- Els investigadors s'han d'adherir als principis de beneficència, no maleficència, autonomia i justícia.
- Altres principis o factors a considerar són: precaució, inclusivitat, responsabilitat, seguretat, reversibilitat, privadesa i maximització de beneficis (i minimització de danys).

2) Llicència per a dissenyadors

- Cal considerar els valors europeus fonamentals.
- Cal introduir principis de disseny fiables per a la seguretat.
- Cal assegurar la privadesa per disseny.
- Cal la presència visible de botons de desconnexió.
- Cal operar el robot d'acord amb principis ètics locals i internacionals.
- Les decisions del robot han de poder ser reconstruïdes.
- Cal màxima transparència i predictibilitat, àdhuc considerant la incertesa.
- Cal assegurar que els robots siguin identificables com a tals.
- Cal obtenir la valoració positiva d'un comitè d'ètica en la recerca abans de provar un robot en un entorn real.

3) Comitè d'Ètica en la Recerca (REC, *Research Ethics Committee*)

- Ha d'ajustar-se als principis d'independència, competència, transparència i responsabilitat.
- S'ha d'establir quina és la funció del REC.
- S'han de proveir els mecanismes de la seva constitució i de la seva labor de monitoratge.

4) Llicència per a l'usuari

- L'usuari ha de poder fer ús d'un robot sense risc o por de patir dany físic o psicològic.
- Té dret a esperar que el robot faci qualsevol tasca per la qual ha estat dissenyat explícitament.
- Ha de tenir consciència de les limitacions perceptuals, cognitives i d'actuació del robot.
- Ha de considerar el respecte a causa de la fragilitat humana, física i psicològica, i les necessitats emocionals.
- Ha de tenir en compte i respectar els drets a la privadesa.
- Està prohibit usar el robot de manera que contravingui principis ètics o legals.
- Està prohibit modificar un robot per utilitzar-lo com a arma.

3.4. Iniciatives normatives i legals en l'àmbit internacional i extracomunitari

3.4.1. Normes ISO/TC 299 Robotics

Els organismes d'estandardització, més àgils o amb una inèrcia menor que l'aparell legislatiu, ja emeten normativa concernent a les noves tecnologies de l'IA i la robòtica. En aquest subapartat ens centrarem en el treball de l'ISO, l'Organització Internacional d'Estandardització (International Organization for Standardization), una ONG que agrupa organitzacions d'estandardització dels 163 estats membres i l'objectiu de la qual és que els productes i serveis que reben la seva acreditació puguin garantir la seguretat, fiabilitat i bona qualitat. A falta d'una legislació específica, el compliment d'aquestes normatives per part d'una empresa pot tenir un cert pes jurídic arribat el cas d'un judici per responsabilitat civil. Aquí, en particular, resumirem el treball del Comitè Tècnic (TC, *Technical Committee*) ISO/TC 299 Robotics.

És natural que fins avui el major esforç normatiu s'hagi concentrat en els robots manipuladors industrials: ja el 1988 es va emetre un TR (*technical report*) que definia alguns termes (ISO/TR 8373:1988 Manipulating industrial robots - Vocabulary), i el 1990 (ISO 9283:1990) es van especificar els criteris de rendiment dels robots manipuladors industrials i els mètodes de comprovació associats. Les normes afecten aspectes genèrics com la seguretat (ISO 10218-1:2011 Robots and robotic devices - Safety requirements for industrial robots - Part 1: Robots, i ISO 10218-2:2011 Robots and robotic devices - Safety requirements for industrial robots - Part 2: Robot systems and integration) o la nomenclatura (a més dels esmentats al principi, ISO 9787:2013 Robots and robotic devices - Coordinate systems and motion nomenclatures, o ISO 9946:1999 Manipulating industrial robots - Presentation of characteristics). També poden fer referència a dispositius concrets, que es refereixen bàsicament als elements

Normatives TR, ISO/TR 8373:1988, ISO 9283:1990

Aquestes normes ja han estat retirades i substituïdes per unes altres; de fet, van ser emeses per un TC diferent i més antic, l'ISO/TC 184 Sistemes d'automatització i integració. Les versions actualment vigents, emeses pel TC 299, són l'ISO 8373:2012 Robots and robotic devices i l'ISO 9283:1998 Manipulating industrial robots - Performance criteria and related test methods respectivament.

o efectors terminals i la seva intercanviabilitat (ISO 9409-1:2004 Manipulating industrial robots - Mechanical interfícies - Part 1: Plates, ISO 9409-2:2004 Idem - Part 2: Shafts, ISO 14539:2000 Manipulating industrial robots - Object handling with grasp-type grippers - Vocabulary and presentation of characteristics, ISO/TR 20218-1:2018 Robotics - Safety design for industrial robot systems - Part 1: End-effectors), o a elements usuals en les cèl·lules robotitzades (ISO/TR 20218-2:2017 Robotics - Safety design for industrial robot systems - Part 2: Manual load/unload stations). Cal destacar, per la proactivitat demostrada a adaptar-se a les noves tendències, l'existència de l'estàndard ISO/TS 15066:2016 Robots and robotic devices - Collaborative robots, relatiu a les normes de seguretat per a robots que (en un entorn industrial) treballen en col·laboració amb persones.

Fora de l'àmbit industrial, com que comença a haver-hi robots mòbils i de serveis que s'ofereixen com a productes comercials, l'ISO també té alguna cosa a dir-hi. La primera norma a veure la llum va ser l'ISO 13482:2014 Robots and robotic devices - Safety requirements for personal care robots¹⁰, que se centra en el robot mòbil servent, el robot d'assistència física i el robot de transport personal, els perills potencials associats i com manejar aquests riscos, alhora que descarta de la seva consideració altres robots com els industrials, els militars, els que es desplacen per aire o per aigua (o per terra a més de 20 km/h), metges i joguines.

La seguretat és la preocupació principal de les normes, i en el cas dels robots d'assistència personals això queda reflectit en la norma ISO 13482:2014 Robots and robotic devices - Safety requirements for personal care robots, els aspectes d'aplicació dels quals s'estan revisant a ISO/PRF TR 23482-1 Robotics - Application of ISO 13482 - Part 1: Safety-related test methods, i ISO/TR 23482-2:2019 Idem - Part 2: Application guidelines. Les Normes ISO 18646 Robotics - Performance criteria and related test methods for service robots regulen diversos aspectes dels robots de serveis: ISO 18646-1:2016 Part 1: Locomotion for wheeled robots, ISO 18646-2:2019 Part 2: Navigation, ISO/CD¹¹ 18646-3 Part 3: Manipulation, i ISO/CD 18646-4 Part 4: Lumbar support robots. També la nomenclatura i la normalització són aspectes que requereixen una estandardització: ISO 19649:2017 Mobile robots - Vocabulary, i ISO/DIS 22166-1 Robotics - Modularity for service robots - Part 1: General requirements.

Finalment, ara com ara la robòtica mèdica també té el seu lloc en les especificacions del comitè: IEC/TR 60601-4-1:2017 Medical electrical equipment - Part 4-1: Guidance and interpretation - Medical electrical equipment and medical electrical systems employing a degree of autonomy, IEC 80601-2-77:2019 Medical electrical equipment - Part 2-77: Particular requirements for the basic safety and essential performance of robotically assisted surgical equipment, i IEC 80601-2-78:2019 Medical electrical equipment - Part 2-78: Particular requirements for basic safety and essential performance of medical robots for rehabilitation, assessment, compensation or alleviation. Tota l'activitat de l'ISO/

Enllaç d'interès

Vegeu també l'article referent a María Lazarte «Robots and humans can work together with new ISO guidance», de 8 de març de 2016. <<https://www.iso.org/news/2016/03/Ref2057.html>>.

⁽¹⁰⁾ Vegeu-ne el resum a <https://www.iso.org/standard/53820.html> i els enllaços a articles de divulgació de la mateixa entitat sobre aquesta normativa en aquesta mateixa pàgina.

⁽¹¹⁾ CD: Committee Draft

IEC

Aquest és l'acrònim d'International Electrotechnical Commission, entitat que col·labora amb l'ISO.

TC 299 està recollida a la pàgina <https://www.iso.org/committee/5915511/x/catalogue/>, i suposem que d'aquí a uns anys tindrà moltes ampliacions i actualitzacions.

3.4.2. Àmbit extracomunitari

El 2017 va saltar als titulars de notícies de tot el món que Riad havia concedit el dret de ciutadania a la robot humanoide Sophia. La notícia sorprenué per dos motius: la irregularitat del procediment (la nacionalitat àrab s'adquireix solament per naixement en una família nativa o per matrimoni amb un nadiu; Sophia tampoc no complia els requisits d'edat, permanència o fluïdesa idiomàtica) i el fet que es garantia a Sophia una autonomia negada a les dones humanes. El cas és analitzat a Atabekov i Yastrebov (2018) juntament amb un altre cas, el de permís de residència atorgat al *chatbot* Shibuya Mirai en un altre país refractari a admetre nous residents, Japó.

Al marge de l'anècdota, que il·lustra l'arbitrarietat dels governants en absència de legislacions clares, s'albiren alguns esforços en països més avançats tecnològicament per a començar a desenvolupar lleis que regulin l'IA i la robòtica:

- **Estats Units.** En un país de tradició jurisprudencial consolidada l'investigador Ryan Calo (2016), professor a la University of Washington School of Law, sosté, després d'examinar casos que involucren robots amb presència a les sales de justícia des dels anys quaranta, que la política legislativa sobre robots no vindrà solament de la jurisprudència sinó de l'esforç legislatiu d'institucions com la National Highway Traffic Safety Administration i la Federal Aviation Association. De fet, ambdues institucions ja van emetre una regulació d'àmbit federal el 2016 en els seus respectius camps (vehicles autònoms i UAV). És a dir, segons Calo, la regulació s'anirà construint *ad hoc* però amb certes pretensions de generalitat en àrees específiques.
- **Regne Unit.** Al marge de les iniciatives de la UE, el 2016 un comitè de la Casa dels Comuns britànica publica l'informe «Robots i IA», amb recomanacions al Govern per a establir una comissió sobre IA (Wright i Bledsoe, 2017).
- **Japó.** El febrer de 2015 l'Oficina Central per a la Revitalització Econòmica del METI (Ministeri Japonès d'Economia, Comerç i Indústria) va emetre la «Nova estratègia dels robots. L'estratègia robòtica del Japó. Visió, estratègia, pla d'acció». La preocupació principal en l'àmbit de l'ètica es refereix als robots assistencials. En aquest informe també s'estableix la Iniciativa de la Revolució Robòtica, una plataforma oberta d'innovació dirigida per la iniciativa privada. El text proporciona recomanacions sense força legal.

«Nova estratègia dels robots. L'estratègia robòtica del Japó. Visió, estratègia, pla d'acció»

http://www.meti.go.jp/english/press/2015/pdf/0123_01b.pdf

- **Xina.** El 2017 el Consell d'Estat va emetre el «Pla de desenvolupament de la nova generació d'IA», on s'estableix l'objectiu de convertir la Xina en la primera potència mundial en IA el 2030 (Karch *et al.*, 2018). Defineix unes «tasques focals» d'aplicació de l'IA a reptes socials, econòmics i de seguretat nacional. També inclou una sèrie de «mesures de garantia» per a guiar «el desenvolupament i aplicació de l'IA, amb lleis i regulacions, marcs ètics i distribució de recursos». El desembre de 2017 el Ministeri d'Indústria i Tecnologia de la Informació va establir un «Pla d'acció trienal per a promoure el desenvolupament d'una nova generació d'indústria de l'IA», basat en el pla anterior, amb especial incidència en «vehicles autònoms, robots intel·ligents de servei, UAV intel·ligents, sistemes de diagnòsi d'imatge mèdica, sistemes d'identificació d'imatge i vídeo, sistemes de veu interactius intel·ligents, sistemes de traducció intel·ligents i productes domòtics».
- **Corea del Sud.** El 2007 el Ministeri de Comerç, Indústria i Energia de Corea anuncia la creació d'una Carta d'ètica en robòtica. No obstant això, en dades més recents la tendència és més aviat a desregularitzar per facilitar la promoció de la tecnologia i la indústria robòtica. Per exemple, s'han relaxat els requisits de seguretat als mínims necessaris per afavorir l'expansió de robots col·laboratius que treballen amb humans a les fàbriques. Els esforços es dirigeixen a «identificar i millorar les regulacions que obstaculitzen la popularització del robot i el desenvolupament de la tecnologia en diversos camps» (Sang-mo, 2018).
- **Rússia.** Des del 2016 la Duma està discutint un esborrany de llei per a regular les relacions entre humans i robots amb vista a la seva aprovació el 2022. Conegut com a «lleï Grishin» pel seu proponent Dmitry Grishin, preveu els casos en què un robot o les seves accions es puguin considerar un crim: la creació d'un robot assassí amb aquesta finalitat, la desactivació dels elements de maquinari i programari que bloquegen la possibilitat de causar dany a éssers humans, la construcció d'un robot capaç de causar dany, i la d'un robot incapaç de *comprendre* que pot ser usat per a causar dany (Goldman, 2017). També es proposa crear un registre de robots.

Iniciativa de la Revolució Robòtica

<https://www.jmfrri.gr.jp/english/outline/763.html>

3.5. Drets del robot: visions de futur

El dret es veurà afectat sens dubte de moltes maneres per l'existència, desenvolupament i difusió dels robots i de l'IA. En el subapartat 3.2 plantegem algunes qüestions sobre responsabilitat civil i penal d'uns robots hipotètics que han evolucionat fins a tenir objectius propis. Si seguim amb aquesta hipòtesi, ens podem preguntar legítimament si el reconeixement de responsabilitat comporta com a contrapartida el reconeixement de drets. El reconeixement de drets d'un subjecte (o conjunt de subjectes) es pot iniciar de dues maneres: el subjecte en qüestió els reclama per a si mateix (per exemple, les dones amb el moviment sufragista a mitjan segle XIX o la comunitat afroamericana

nord-americana als anys cinquanta del passat segle) o bé són reclamats per qui pot fer-ho (per exemple, els qui advoquen pels drets dels animals). La primera via sembla implicar alguna manera de verbalitzar un desig, una aspiració per a fer-se efectiva, i requereix, per a prosperar, que qui pot atorgar els drets reclamats admeti que el col·lectiu en qüestió és qualificat per a gaudir-los. Al seu moment això va comportar que un cert col·lectiu d'homes blancs admetessin que no hi havia diferències essencials entre ells i el col·lectiu de dones o de les persones de color i que la igualtat de drets estava justificada. En casos que requereixen la segona via no es pretén generalment una equiparació de drets, però sí garantir una sèrie de drets bàsics, com el dret a la vida o a no ser torturat. Els robots són ara com ara màquines que encara disten molt de poder-se considerar per al reconeixement de drets semblants, si més no, als d'algunes categories d'animals. Tenen el mateix dret a no ser desballestats que una rentadora. La verbalització no comporta cap problema per als robots a diferència dels animals: poden sintetitzar perfectament la frase «Vull que se'm reconegui el dret a no ser desballestat». Una cosa diferent és què hi ha darrere de la primera paraula d'aquesta frase. No es pot afirmar (encara?, mai?) que els robots tinguin capacitat de sentir, a diferència dels animals. No fa falta arribar a una consciència del mateix nivell que la humana: per a ser candidat al dret de no-desballestament, probablement n'hi hauria prou amb «sentir, percebre o experimentar subjectivament» (Wikipedia, «Sintiencia», 2019). És la capacitat de sentir una característica exclusiva dels éssers biològics? Dels éssers amb un organisme basat en la química del carboni? No obstant això, tampoc no es pot obviar que darrere de la frase en qüestió pot ser que no hi hagi un desig (subjectiu o no) del robot que la pronuncia sinó de l'empresa que el va fabricar o del seu propietari com a via per a expandir el seu propi poder en el si de la societat mitjançant els seus robots.

Sembla que encara és prematur plantejar-se aquestes qüestions. No obstant això, la controvèrsia i el debat són aquí. La professora Joanna Bryson, de la Universitat de Bath, sosté que, encara que els robots es puguin arribar a (qualificar per a) tenir drets, no haurien de tenir-los: «Els robots haurien de ser esclaus». Quin sentit té crear-los si no? Ja ho són etimològicament: servents que fan treballs forçats. Però, tal com exposa David J. Gunkel en el seu llibre *Robots rights*, no és solament la qüestió que retornar a una societat esclavista podria comprometre certament el nostre sentit de l'ètica, sinó que les societats esclavistes del passat tenien sistemes legals complexos que incloïen alguns drets dels esclaus. A això es podria afegir que, si és produït l'emergència d'aspiracions de «llibertat» en el col·lectiu esclavitzat, aquest tindria sens dubte una legitimació moral per a rebel·lar-se. No hi ha garanties que es pugui evitar que en el futur una IA prou complexa i immersa en el món arribi a desenvolupar subjectivitat, consciència, emocions o alguna cosa que se li assembli. La necessitat o inevitabilitat que això passi és qüestionable, però interrompre tota recerca en IA perquè no pugui passar en cap cas tampoc no sembla molt factible.

No es pot obviar tampoc la possibilitat que els humans acabem fonent-nos, hibridant amb les nostres criatures artificials, tal com preconitza el transhumanisme. De fet, és una possibilitat bastant plausible davant la promesa de capacitats millorades, una experiència vital en formes que avui ens resulten difícils d'imaginar, i una major longevitat de qualitat. On acabarà l'ésser humà i on començarà el robot? Les alteracions genètiques i la nanotecnologia poden difuminar la frontera. ¿Aquestes criatures veuran els seus drets humans minvats en proporció a la seva part artificial? No té massa sentit que sigui així.

Un robot que doni cos a una IA general immersa en el món i avançada, com qualsevol de nosaltres, atresorarà un compendi d'informació únic i irrepetible. A diferència de nosaltres, podrà transferir-se al ciberespai i trobar-hi el seu santuari. O potser no, potser el seu cos, marcat per mil experiències com el nostre, serà indissociable de la intel·ligència que conté. Destruir-lo, esborrar aquesta informació de l'univers, potser tindrà la mateixa transcendència que la mort de qualsevol persona, és a dir, molt poca relativament. Però al mateix temps alguna cosa molt valuosa s'estaria perdent. Per tant, no sembla desgavellat que almenys en la petita escala de la societat humana s'intenti protegir aquesta existència mitjançant el dret a ella.

Resum

La robòtica i l'IA tenen un potencial transformador de la societat humana que a llarg termini poden fer-la difícilment recognoscible. Solament és comparable a les possibilitats que ofereix la manipulació genètica. El que ens pot oferir en el futur llunyà és una incògnita. Però en un termini més breu sí que podem anticipar, avaluar i intentar dirigir certes tendències. Per a això, hem de tenir una imatge total de què estem parlant, què és el que ja poden fer els robots, com s'està utilitzant l'IA i quines són les possibilitats reals d'evolució. D'això s'ha ocupat la primera part, on hem vist que els robots poden donar un cos a l'IA amb el qual poden percebre el món i actuar-hi amb un grau de sofisticació i versatilitat que excedeix el de les màquines convencionals. Les capacitats d'aprenentatge, raonament, representació del coneixement, presa de decisions i elaboració de plans d'acció complexos són les eines d'IA que permeten al robot mostrar un comportament intel·ligent. D'aquesta manera, el robot pot anar penetrant en molts àmbits de l'activitat humana, encara que durant molt temps encara li faltará la capacitat d'adaptació de l'ésser humà a entorns i circumstàncies molt dispars, canviants i ambigües. Durant molt temps no tindrà (si arriba a tenir algun dia) sentit comú, ni consciència ni emocions, i per tant serà difícil que puguem atribuir-li capacitat de sentir. Malgrat això, exercirà un paper cada vegada més important en la societat, i el seu contacte amb els éssers humans serà cada vegada major, més complex i ens obligarà a plantejar-nos i intentar resoldre els reptes de caràcter ètic que s'han descrit en la segona part. La resolució dels conflictes que emergeixin requerirà dilucidar qüestions relacionades amb la responsabilitat civil de les accions del robot. Aquest és el sentit primordial dels tímids esforços normatius i legislatius que es comencen a produir, que també hauran d'incidir en altres aspectes de l'ètica no relacionats amb la seguretat física però sí amb la dignitat i la integritat psicològica i emocional de les persones, tal com s'ha vist en la tercera part.

Recomanacions bibliogràfiques:

Donada la complexitat i la vasta importància de l'IA i la robòtica en els diferents aspectes de la vida humana i de la societat, aquí solament hem pogut presentar-ne una imatge simplificada i general. Podem recomanar les lectures següents (que són en les referències esmentades al llarg del text):

López de Mántaras Badía, R.; Meseguer González, P. (2017). *Inteligencia artificial*. CSIC / Los Libros de la Catarata («¿Qué sabemos de?»).

Lin, P.; Abney, K.; Bekey G .A. (eds.) (2014). *Robot ethics. The ethical and social implications of robotics*. The MIT Press.

Lin, P.; Jenkins, R.; Abney, K. (eds.) (2017). *Robot ethics 2.0. From autonomous cars to artificial intelligence*. Oxford University Press.

Torras, C. (2018) *The vestigial heart*. A novel of the robot age, with an appendix discussing 24 ethics questions, and associated online instructor resources. The MIT Press¹².

(12) Es tracta d'una obra de ficció, però precisament per això té un gran valor educatiu, tal com s'indica a Torras (2019) citant Burton *et al.* (2018).

Per a estar al dia sobre l'evolució del dret relacionat amb l'IA, es poden consultar els enllaços següents:

Law and AI. «A blog devoted to examining the law of artificial intelligence, AI in law, and AI policy». <http://www.lawandai.com/>

International Association for Artificial Intelligence and Law, <http://www.iaail.org/>, que organitza anualment la International Conference of Artificial Intelligence and Law (ICAAIL).

We Robot Conference on Legal and Policy Issues Relating to Robotics. Edició de 2020. <http://robots.law.miami.edu/>

Bibliografia

Allen, C.; Wallach, W. (2014). «Moral machines: contradiction in terms or abdication of human responsibility?». A: K. Abney; P. Lin; G. A. Bekey (eds.). *Robot ethics. The ethical and social implications of robotics* (pàg. 57). The MIT Press.

Arkin, R. (2009). *Governing lethal behavior in autonomous robots*. Boca Raton, FL: Chapman and Hall / CRC Press.

Asaro, P. M. (2014). «A body to kick, but still no soul to damn: legal perspectives on robotics». A: K. Abney; P. Lin; G. A. Bekey (eds.). *Robot ethics: the ethical and social implications of robotics*. MIT Press.

Atabekov, A.; Yastrebov, O. (2018). «Legal status of artificial intelligence across countries: legislation on the move». *European Research Studies Journal* (vol. XXI, núm. 4, pàg. 773-782).

Awad, E.; Dsouza, S.; Kim, R. et al. (2018). «The moral machine experiment». *Nature* (vol. 563, pàg. 59-64).

Borenstein, J.; Howard, A.; Wagner, A. R. (2017). «Pediatric robotics and ethics». A: K. Abney; P. Lin; G. A. Bekey (eds.). *Robot ethics 2.0. From autonomous cars to artificial intelligence* (pàg. 127-141). Oxford University Press.

Brown, J. (2019). «Tesla autopilot malfunction caused crash that killed Apple engineer, lawsuit alleges». *Gizmodo*. <<https://gizmodo.com/tesla-autopilot-malfunction-caused-crash-that-killed-ap-1834453661>>.

Buchanan, B. G.; Headrick, T. E. (1970). «Some speculation about artificial intelligence and legal reasoning». *Stanford Law Review* (pàg. 40-62).

Burton, E.; Goldsmith, J.; Mattei, N. (2018). «How to teach computer ethics through science fiction». *Communications of the ACM* (vol. 61, núm. 8, pàg. 54-64).

Calo, R. (2014). «Robots and privacy». A: K. Abney; P. Lin; G. A. Bekey (eds.). *Robot ethics. The ethical and social implications of robotics* (pàg. 187-201). The MIT Press.

Calo, R. (2016). «Robots in American law» (Legal Studies Research Paper No. 2016-04). University of Washington School of Law. <<http://euro.ecom.cmu.edu/program/law/08-732/AI/Calo.pdf>>.

Cargill (2018). «Meet the robot that's making cattle herding safer». <<https://www.cargill.com/story/meet-the-cowboy-robot-thats-making-cattle-herding-safer>>.

Chakravorty, A. (2019). «Underground robots: how robotics is changing the mining industry». *EOS, Earth and Space Science News* (13 de maig de 2019). <<https://eos.org/features/underground-robots-how-robotics-is-changing-the-mining-industry>>.

Chen, F.; Favetto, A.; Mousavi, M. et al. (2011). «Human hand: kinematics, statics and dynamics». *International Conference on Environmental Systems, July 17-21, 2011*. Portland, Oregon. <https://pdfs.semanticscholar.org/2498/fc5ce655c7085a330657e2a281a1e75ced71.pdf?_ga=2.16791225.1216372352.1569997220-974663121.1563870901>.

Da Vinci (2019). *Intuitive*. Robotic assisted systems, Da Vinci robot. <<https://www.intuitive.com/en-us/products-and-services/da-vinci/systems>>. [Data de consulta: 30 de setembre de 2019].

Elder, A. (2017). «Robot friends for autistic children. Monopoly money or counterfeit currency?». A: K. Abney; P. Lin; F. A. Bekey (eds.). *Robot ethics 2.0. From autonomous cars to artificial intelligence* (pàg. 113-126). Oxford University Press.

European Commission (2014). «Robots help each other in RoboEarth». *Digital Single Market* (16 de gener de 2014). <<https://ec.europa.eu/digital-single-market/en/news/robots-help-each-other-roboearth>>.

Foot, P. (1978). *The problem of abortion and the doctrine of the double effect in virtues and vices*. Oxford: Basil Blackwell.

Frey, C. B.; Osborne, M. A. (2013). «The future of employment: how susceptible are jobs to computerisation?». *University of Oxford*.

Fukuoka (2004). «Robot World Declaration». International Robot Fair 2004. <<http://prw.kyodonews.jp/prwfile/prdata/0370/release/200402259634/index.html>>.

Gendron, J. (2019). «How mining robots are replacing humans and saving lives». *RobotShop Community* (10 de juliol de 2019). <<https://www.robotshop.com/community/blog/show/how-mining->+i+robots-are-replacing-humans-and-saving-lives>>.

Goldman, E. (2017). «Russian government plans robot laws». *Russia Beyond* (14 de febrer de 2017). <https://www.rbth.com/science_and_tech/2017/02/14/russian-government-plans-robot-laws_701886>.

González, M. (2015). «Los héroes robóticos de Fukushima». *Xataka* (13 de març de 2015). <<https://www.xataka.com/robotica-e-ia/los-heroes-roboticos-de-fukushima>>.

Guarini, M.; Bello, P. (2014). «Robotic warfare: some challenges in moving from noncivilian to civilian theaters». A: K. Abney; P. Lin; G. A. Bekey (eds.). *Robot ethics. The ethical and social implications of robotics* (pàg. 132). The MIT Press.

Jiménez de la Fuente, F. (2019). «Me casé con un holograma: es difícil de entender, pero debería ser respetado». *BBC Mundo* (28 de maig de 2019). <<https://www.bbc.com/mundo/noticias-internacional-48229491>>.

Jiménez, P. (2018). «Software tools for the cognitive development of autonomous robots». *Autonomy in future military and security technologies: implications for law, peace, and conflict* (pàg. 18-32). Richardson Institute / Lancaster University. <<http://www.iri.upc.edu/files/sci-doc/2106-Software-tools-for-the-cognitive-development-of-autonomous-robots.pdf>>.

Kahn, L. (2017). «Military Robots and the Likelihood of Armed Combat», En: Abney, K.; Lin, P.; Bekey, G. A. (eds.) *Robot Ethics 2.0. From Autonomous Cars to Artificial Intelligence*. Oxford University Press (pàg. 274-287).

Karch, T. J.; Kaja, A.; Luo, Y. (2018). «China's vision for the next generation of artificial intelligence». *The National Law Review* (25 de març de 2018). <<https://www.natlawreview.com/article/china-s-vision-next-generation-artificial-intelligence>>.

Kirkpatrick, J.; Hahn, E. N.; Haufler, A. J. (2017). «Trust and Human-Robot Interactions», En: Abney, K.; Lin, P.; Bekey, G. A. (eds.): *Robot Ethics 2.0. From Autonomous Cars to Artificial Intelligence*. Oxford University Press (pàg. 142-156).

Levy, D. (2014). «The ethics of robot prostitutes». A: K. Abney; P. Lin; G. A. Bekey (eds.). *Robot ethics. The ethical and social implications of robotics* (pàg. 223-231). The MIT Press.

Loh, W.; Loh, J. (2017). «Autonomy and responsibility in hybrid systems». A: K. Abney; P. Lin; G. A. Bekey (eds.). *Robot ethics 2.0. From autonomous cars to artificial intelligence* (pàg. 35-50). Oxford University Press.

Lokhorst, G.-J.; Van den Hoven, J. (2014). «Responsibility for military robots». A: K. Abney; P. Lin; G. A. Bekey (eds.). *Robot ethics. The ethical and social implications of robotics* (pàg. 145-156). The MIT Press.

López de Mántaras Badía, R.; Meseguer González, P. (2017). «Inteligencia artificial». CSIC / Los Libros de la Catarata («¿Qué sabemos de?»).

Mars Exploration Program (2019). <<https://mars.jpl.nasa.gov>> [Data de consulta: 30 de setembre de 2019].

Meacham, D.; Studley, M. (2017). «Could a robot care? It's all in the movement». A: K. Abney; P. Lin; G. A. Bekey (eds.). *Robot ethics 2.0. From autonomous cars to artificial intelligence* (pàg. 93-112). Oxford University Press.

Michigan State University (2019). «Robotic Milking». *W. K. Kellogg Farm* [en línia]. <https://farm.kbs.msu.edu/pdc/robotic-milking/>. [Data de consulta: 30 de setembre de 2019].

Millar, J. (2014). «An ethical dilemma: when robot cars must kill, who should pick the victim?». *Robohub* (11 de juny de 2014). <<https://robohub.org/an-ethical-dilemma-when-robot-cars-must-kill-who-should-pick-the-victim/>>.

Millar, J. (2017). «Ethics settings for autonomous vehicles». A: K. Abney; P. Lin; G. A. Bekey (eds.). *Robot ethics 2.0. From autonomous cars to artificial intelligence* (pàg. 20-34). Oxford University Press.

Montañez, G. (2019). «AI: think about ethics before trouble arises». *Mind Matters News* (7 de gener de 2019). <<https://mindmatters.ai/2019/01/ai-think-about-ethics-before-trouble-arises/>>.

Moor, J. (2006). «The nature, importance and difficulty of machine ethics». *IEEE Intelligent Systems* (vol. 21, núm. 4, pàg. 18-21).

Muscato, G.; Bonaccorso, F.; Cantelli, L. et al. (2012). «Volcanic environments. Robots for exploration and measurement». *IEEE Robotics & Automation Magazine* (març de 2012, pàg. 40-49). <https://www.academia.edu/10216591/Volcanic_Environments_Robots_for_Exploration_and_Measurement>.

Noticias ONU (2019). «El cambio climático costará 80 millones de puestos de trabajo en 2030» (1 de juliol de 2019). <<https://news.un.org/es/story/2019/07/1458652>>.

O'Leary, D. (2019). «Can an algorithm be racist?». *Mind Matters News* (3 de gener de 2019). <<https://mindmatters.ai/2019/01/can-an-algorithm-be-racist/>>.

O'Meara, R. M. (2014). «Contemporary governance architecture regarding robotics technologies: an assessment». A: K. Abney; P. Lin; G. A. Bekey (eds.). *Robot ethics. The ethical and social implications of robotics* (pàg. 159-168). The MIT Press.

O'Neil, C. (2016). *Weapons of math destruction: how big data increases inequality and threatens democracy*. Crown Random House.

Open Roboethics Initiative (2014). «My (autonomous) car, my safety: results from our reader poll». *Robohub* (30 de juny de 2014). <<https://robohub.org/my-autonomous-car-my-safety-results-from-our-reader-poll/>>.

Oxford Economics (2019). «Report - How robots change the world» (juny de 2019). <<http://resources.oxfordeconomics.com/how-robots-change-the-world>>.

Pierson, A. (2015). «Bio-inspired non-cooperative multi-robot herding». <<https://alyssapierson.com/research/herding/>>.

Postscapes (2019). «Agriculture robots» (25 de març de 2019). <<https://www.postscapes.com/agriculture-robots/>>.

Robotonomics (2015). «Study-robots are not taking jobs». <<https://www.the-vital-edge.com/robotics-and-job-loss-wait-not-so-fast/>> [L'enllaç a l'article original no funcionava a data de 2 d'octubre de 2019].

Robolaw Asia (2012). «Robolaw.Asia». YSAiL Initiative for Robotics, Law & Policy. <<http://www.robolaw.asia/index-pku.html>>.

Robolaw (2014). «Regulating robotics: a challenge for Europe». <[http://www.europarl.europa.eu/RegData/etudes/IDAN/2014/509987/IPOL_IDA\(2014\)509987\(ANN01\)_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/IDAN/2014/509987/IPOL_IDA(2014)509987(ANN01)_EN.pdf)>.

Robotic Industries Association (2019). «A tribute to Joseph Engelberger. Unimate, the first industrial robot». <<https://www.robotics.org/joseph-engelberger/unimate.cfm>>. Acceso: 30 de septiembre de 2019

Robotics Online Marketing Team (2018). «Construction robots will change the industry forever». *Robotics Online Blog* (17 d'abril de 2018). <<https://www.robotics.org/blog-article.cfm/Construction-Robots-Will-Change-the-Industry-Forever/93>>.

SAE International (2019). «SAE J3016TM, levels of driving automation». <<https://www.sae.org/news/2019/01/sae-updates-j3016-automated-driving-graphic>>.

Román Maestre, B. (2016). *Ética de los servicios sociales*. Madrid: Herder.

Sang-mo, K. (2018). «Policy directions for S. Korea's robot industry». *BussinessKorea* (17 d'agost de 2018). <<http://www.businesskorea.co.kr/news/articleView.html?idxno=24394>>.

Scheutz, M. (2014). «The inherent dangers of unidirectional emotional bonds between humans and social robots». A: K. Abney; P. Lin; G. A. Bekey (eds.) *Robot ethics. The ethical and social implications of robotics* (pàg. 204-221). The MIT Press.

Shachtman, N. (2007). «Robot cannon kills 9, wounds 14». *Wired* (18 d'octubre de 2007). <<https://www.wired.com/2007/10/robot-cannon-ki/>>.

Sharkey, N. (2014). «Killing made easy». A: K. Abney; P. Lin; G. A. Bekey (eds.). *Robot ethics. The ethical and social implications of robotics* (pàg. 111-128). The MIT Press.

Smithsonian Ocean Team (2009). «From submarines to robots: exploring the deep ocean». <<https://ocean.si.edu/ecosystems/deep-sea/submarines-robots-exploring-deep-ocean>>.

Surden, H. (2019). «Artificial Intelligence and Law: An Overview». *Georgia State University Law Review* (vol. 35, pàg. 1305-1337). U. of Colorado Law Legal Studies Research Paper No. 19-22, <<https://ssrn.com/abstract=3411869>>.

Torrás, C. (2016). «Service robots for citizens of the future». *European Review* (vol. 24, núm. 1, pàg. 17-30). <<http://www.iri.upc.edu/files/scidoc/1719-Service-Robots-for-Citizens-of-the-Future.pdf>>.

Torrás, C. (2016b). «Robot pain: a speculative review of its functions». *Pain and the conscious brain* (pàg. 235-246). Wolters Kluwer. <<http://www.iri.upc.edu/files/scidoc/1755-Robot-pain:-a-speculative-review-of-its-functions.pdf>>.

Torrás, C. (2018). *The vestigial heart*. Josephine Swarbrick (trad.). MIT Press. <<http://www.iri.upc.edu/people/torras/vestigial.html>>. Versió en castellà: *La mutación sentimental* (2012). Ed. Milenio. Versió en català: *La mutació sentimental* (2008). Pagès Editors.

Torrás, C. (2018b). «Social networks and robot companions: technology, ethics and science fiction». *Mètode Science Studies Journal* (vol. 99, pàg. 47-53). <http://www.iri.upc.edu/files/scidoc/2096-SOCIAL->_i_NETWORKS-AND-ROBOT-COMPANIONS:-TECHNOLOGY,-ETHICS,-AND-SCIENCE-FICTION.pdf>.

Torrás, C. (2019). «Assistive robotics: research challenges and ethics education initiatives». *DILEMATA: International Journal of Applied Ethics* (vol. 30, pàg. 63-77). <<https://www.dilemata.net/revista/index.php/dilemata/article/view/412000291/641>>.

Wallach, W.; Allen, C. (2009). *Moral machines. Teaching robots right from wrong*. Nova York: Oxford University Press.

Whitby, B. (2014). «Do you want a robot lover?». A: K. Abney; P. Lin; G. A. Bekey (eds.). *Robot ethics. The ethical and social implications of robotics* (pàg. 233-248). The MIT Press.

White, T. N.; Baum, S. D. (2017). «Liability for present and future robotics technology». A: K. Abney; P. Lin; G. A. Bekey (eds.). *Robot ethics 2.0. From autonomous cars to artificial intelligence* (pàg. 66-79). Oxford University Press.

Wikipedia (2019). «Kenji Urada». <https://en.wikipedia.org/wiki/Kenji_Urada>. [Data de consulta: 2 d'octubre de 2019].

Wikipedia (2019). «Legal expert system». <https://en.wikipedia.org/wiki/Legal_expert_system>. [Data de consulta: 2 d'octubre de 2019].

Wikipedia (2019). «Legal Informatics». <https://en.wikipedia.org/wiki/Legal_informatics>. [Data de consulta: 2 d'octubre de 2019].

Wikipedia (2019). «Robert Williams». <[https://en.wikipedia.org/wiki/Robert_Williams_\(robot_fatality\)](https://en.wikipedia.org/wiki/Robert_Williams_(robot_fatality))>. [Data de consulta: 2 d'octubre de 2019].

Wikipedia (2019). «Sintiencia». <<https://es.wikipedia.org/wiki/Sintiencia>>. [Data de consulta: 2 d'octubre de 2019].

Wikipedia (2019). «Virtue ethics». <https://en.wikipedia.org/wiki/Virtue_ethics>. [Data de consulta: 2 d'octubre de 2019].

Winfield, A. (2018). «A roundup of robotics and AI ethics: part 1 principles». *Robohub* (8 de gener de 2018). <<https://robohub.org/a-round-up-of-robotics-and-ai-ethics-part-1-principles/>>.

World Robotics (2019). «Preview». <https://ifr.org/downloads/press2018/IFR_World_Robotics_Outlook_2019_-_Chicago.pdf>.

World Robotics (2018). «Executive Summary World Robotics 2018 Industrial Robots». <https://ifr.org/downloads/press2018/Executive_Summary_WR_2018_Industrial_Robots.pdf>.

Wright, T.; Bledsoe, K. D. (2017). «Robots and rule-makers». *Lexology* (7 de març de 2017). <<https://www.lexology.com/library/detail.aspx?g=07fd6342-104a-4212-8d40-5ecf71f28800>>.

Yan, A. (2017). «Chinese robot dentist is first to fit implants in patient's mouth without any human involvement». *South China Morning Post* (21 de setembre de 2017). <<https://www.scmp.com/news/china/article/2112197/chinese-robot-dentist-first-fit-implants-patients-mouth-without-any-human>>.

