Passat, present i futur de la intel·ligència artificial: algunes reflexions

Discurs de presentació de Ramon López de Mántaras i Badia com a membre numerari de la Secció de Ciències i Tecnologia, llegit el dia 19 de febrer de 2018

Institut d'Estudis Catalans

SECCIÓ DE CIÈNCIES I TECNOLOGIA

Passat, present i futur de la intel·ligència artificial: algunes reflexions

Passat, present i futur de la intel·ligència artificial: algunes reflexions

Discurs de presentació de Ramon López de Mántaras i Badia com a membre numerari de la Secció de Ciències i Tecnologia, llegit el dia 19 de febrer de 2018

Barcelona, 2018

Institut d'Estudis Catalans

SECCIÓ DE CIÈNCIES I TECNOLOGIA

Biblioteca de Catalunya. Dades CIP

López de Mántaras, Ramon, 1952- autor
Passat, present i futur de la intel·ligència artificial : algunes reflexions
Bibliografia
ISBN 9788499654003
I. Institut d'Estudis Catalans. Secció de Ciències i Tecnologia II. Títol
1. Intel·ligència artificial
004.8

Carrer del Carme, 47. 08001 Barcelona

Primera edició: febrer del 2018

Text revisat lingüísticament per la Unitat de Correcció del Servei Editorial de l'IEC

Disseny de la coberta: Azcunce | Ventura

Compost per Accent Llibres, SL
Imprès a QP Print, SL

ISBN: 978-84-9965-400-3
Dipòsit Legal: B 824-2018

Introducció

És possible construir màquines intel·ligents? És una màquina, el cervell? Aquestes són dues preguntes que han obsessionat grans pensadors durant segles. El desenvolupament de la intel·ligència artificial (IA) ha acostat les dues preguntes i fins i tot, per a molts investigadors, les ha unificades en el sentit que s'estan fent servir conceptes, tècniques i experiments semblants en els intents de dissenyar màquines intel·ligents i investigar la naturalesa de la ment. Actualment sabem encara relativament poc sobre el cervell, però estem seguint un camí que implica considerar-lo un sistema computacional i hem començat a explorar l'espai de possibles models computacionals que permetin emular-ne el funcionament.

L'objectiu últim de la IA, aconseguir que una màquina tingui una intel·ligència de tipus general similar a la humana, és un dels objectius més ambiciosos que s'ha plantejat la ciència. Per la seva dificultat, és comparable a altres grans objectius científics com explicar l'origen de la vida, l'origen de l'univers o conèixer l'estructura de la matèria. Al llarg dels darrers segles, aquest afany per construir màquines intel·ligents ens ha conduït a inventar models o metàfores del cervell humà. Per exemple, al segle XVII, Descartes es va preguntar si un complex sistema mecànic compost d'engranatges, politges i tubs podria, en principi, emular el pensament. Dos segles després, la metàfora fou els sistemes telefònics, ja que semblava que les seves connexions es podien assimilar a una xarxa neuronal. Actualment, el model dominant és el model computacional basat en l'ordinador digital i, per tant, és el model al qual em referiré ara i aquí.

La hipòtesi del sistema de símbols físics: IA feble *versus* IA forta

En una ponència, amb motiu de la recepció del prestigiós premi Turing l'any 1975, Allen Newell i Herbert Simon (Newell i Simon, 1976) van formular la hipòtesi del sistema de símbols físics, segons la qual tot sistema de símbols físics posseeix els mitjans necessaris i suficients per dur a terme accions intel·ligents. D'altra banda, atès que els éssers humans som capaços de mostrar conductes intel·ligents en el sentit general, d'acord amb la hipòtesi, nosaltres també som sistemes de símbols físics. Convé aclarir a què es refereixen Newell i Simon quan parlen de sistema de símbols físics (SSF). Un SSF consisteix en un conjunt d'entitats anomenades *símbols*, que, mitjançant relacions, poden ser combinades formant estructures més grans —com els àtoms que es combinen formant molècules— i que poden ser transformades aplicant un conjunt de procediments. Aquests procediments poden crear nous símbols, crear i modificar relacions entre símbols, emmagatzemar símbols, comparar si dos símbols són iguals o diferents, etcètera. Aquests símbols són físics, en tant que tenen un substrat físic i electrònic (en el cas dels ordinadors) o físic i biològic (en el cas dels éssers humans). Efectivament, en el cas dels ordinadors, els símbols es realitzen mitjançant circuits electrònics digitals i en el cas dels éssers humans, mitjançant xarxes de neurones. En definitiva, d'acord amb la hipòtesi SSF, la naturalesa del substrat (circuits electrònics o xarxes neuronals) no té importància, sempre que aquest substrat permeti processar símbols. No oblidem que es tracta d'una hipòtesi i, per tant, no ha de ser ni acceptada ni rebutjada *a priori*. En qualsevol cas, la seva validesa o refutació s'haurà de verificar, d'acord amb el mètode científic, amb assajos experimentals. La IA és precisament el camp científic dedicat a intentar verificar aquesta hipòtesi en el context dels ordinadors digitals, és a dir, verificar si un ordinador convenientment programat és capaç o no de tenir conducta intel·ligent de tipus general.

És important el matís que s'hauria de tractar d'intel·ligència de tipus general i no d'una intel·ligència específica, ja que la intel·ligència dels éssers humans és de tipus general. Exhibir intel·ligència específica és una altra cosa ben diferent. Per exemple, els programes que juguen als escacs al nivell de Gran Mestre són incapaços de jugar a les dames, tot i ser un joc molt més senzill. Es requereix dissenyar i executar un programa diferent i independent per tal que el mateix ordinador jugui a les dames, és a dir, que no pot aprofitar el fet que juga als escacs per adaptar-se i jugar també a les dames. En el cas dels éssers humans no és així, ja que qualsevol jugador d'escacs pot aprofitar els seus coneixements sobre aquest joc per jugar a les dames perfectament en qüestió de minuts o fins i tot de segons. La intel·ligència artificial que únicament mostra comportament intel·ligent en un àmbit molt específic està relacionada amb el que es coneix com a *IA feble* en contraposició amb la *IA forta*, a la qual, de fet, es referien Newell i Simon i altres pares

fundadors de la IA. Encara que estrictament la hipòtesi SSF es va formular el 1975, ja estava implícita en les idees dels pioners de la IA en els anys cinquanta.

Qui va introduir aquesta distinció entre IA feble i IA forta fou el filòsof John Searle en un article crític amb la IA publicat el 1980 (Searle, 1980) que va provocar, i segueix provocant, molta polèmica. La IA forta implicaria que un ordinador convenientment programat no simula una ment, sinó que és una ment, i, per tant, hauria de ser capaç de pensar igual que un ésser humà. Searle, en el seu article, intenta demostrar que la IA forta és impossible. En aquest punt convé aclarir que no és el mateix IA general que IA forta. Hi ha, òbviament, una connexió, però només en un sentit, és a dir, que tota IA forta serà necessàriament general, però hi pot haver IA generals que no siguin fortes, és a dir, que simulin la capacitat d'exhibir intel·ligència general de la ment, però sense ser ments.

La IA feble, d'altra banda, consistiria, segons Searle, a construir programes que realitzin tasques específiques. La capacitat dels ordinadors per portar a terme tasques específiques fins i tot millor que les persones ja s'ha demostrat àmpliament. En certs dominis, els avenços de la IA específica superen molt la perícia humana, com ara buscar solucions a fórmules lògiques amb moltes variables o jugar a escacs, o al Go, o en diagnòstic mèdic i altres aspectes relacionats amb la presa de decisions. També s'associa amb la IA feble el fet de formular i provar hipòtesis sobre aspectes relacionats amb la ment (per exemple, la capacitat de raonar deductivament, d'aprendre inductivament, etc.) mitjançant la construcció de programes que duen a terme aquestes funcions, encara que sigui amb processos completament diferents dels que utilitza el cervell. Absolutament tots els avenços aconseguits fins ara en el camp de la IA són manifestacions de la IA feble i específica.

Turing i la intel·ligència artificial

La idea de la IA forta ja era present fins i tot en els escrits pioners d'Alan Turing (Turing, 1948, 1950) sobre màquines intel·ligents. Alan Turing no tan sols va establir els fonaments teòrics de la computació amb el que ara es coneix com a *màquina de Turing*, sinó que també se'l considera el pare de la intel·ligència artificial. En un article publicat l'any 1950, Turing argumentava que al cap d'uns cinquanta anys hi hauria ordinadors intel·ligents amb la capacitat de raonar deductivament, aprendre nous coneixements per inducció i per experiència i evolució i capaços de comunicar mitjançant interfícies humanitzades.

Turing estava molt interessat en el funcionament del cervell. Estava convençut que el còrtex cerebral d'un nen petit podia ser simulat mitjançant un ordinador. El 1948 va escriure sobre això en un report intern i poc conegut del National Physical Laboratory, titulat «Intelligent machinery», i, en fer-ho, va descriure el que

ara coneixem amb el nom de *xarxes neuronals artificials*. L'esmentat article, que de fet no es va publicar fins a l'any 1969 com un dels capítols del llibre *Machine intelligence 5*, editat per Bernard Meltzer i Donald Michie, presenta un model del cervell basat en unitats de processament molt senzilles. Aquestes unitats estan connectades entre si de manera aleatòria formant una xarxa. Els senyals que processen són binaris, per això actualment anomenaríem aquestes xarxes, *xarxes booleanes*. Turing les va anomenar *màquines no organitzades de tipus A* i la sortida de cada unitat es calcula mitjançant el producte dels valors binaris de les entrades restat d'1, és a dir, que, de fet, cada unitat és una porta lògica de tipus NAND. Aquestes màquines de tipus A no podien aprendre, de manera que Turing les va ampliar afegint una espècie d'interruptors en les connexions entre les neurones que podien ser entrenats per un agent extern que ensenyaria a la màquina a resoldre una tasca determinada. Turing les va anomenar *màquines de tipus B*. L'entrenament de la xarxa consistiria a bloquejar o desbloquejar la connexió entre neurones mitjançant l'interruptor fins a arribar a una màquina «convenientment organitzada», per dur a terme la tasca per a la qual ha estat entrenada. No obstant això, Turing no va proposar cap algoritme per entrenar-la. Poc després de la seva mort es va poder demostrar que aquestes xarxes neuronals booleanes, efectivament, es poden entrenar per aprendre a discriminar entre classes linealment separables. Actualment, xarxes neuronals artificials, organitzades per capes, molt més complexes que les proposades per Turing, s'usen extensivament en IA i robòtica. En aquest mateix treball, Turing també descriu unes màquines que ell anomena *de tipus P*, que es podrien entrenar mitjançant un procés de «premi o càstig», és a dir, que de nou Turing va anticipar el que actualment es coneix com *aprenentatge per reforç*, una de les tècniques d'aprenentatge més reeixides en IA. Les màquines de tipus P, contràriament a les de tipus A i B, no eren xarxes neuronals binàries, sinó màquines de Turing modificades, de manera que, abans de ser entrenades, el conjunt de les seves regles internes és incomplet, però després de ser entrenades contindrien un conjunt complet de regles.

En el context d'aquests treballs sobre màquines intel·ligents, Turing no podia passar per alt la qüestió de com esbrinar si una màquina és intel·ligent o no i, per tal de respondre aquesta pregunta, va proposar el test que porta el seu nom, en un article publicat a la revista *Mind* l'any 1950 i titulat «Computing machinery and intelligence» (Turing, 1950). El test de Turing és una variant de l'anomenat *joc d'imitació*, en el qual, en la versió original, participaven tres persones: un interrogador, un home i una dona. L'interrogador se situa en una sala diferent i es comunica amb les altres dues persones mitjançant missatges de text usant un terminal d'ordinador i disposa de cinc minuts per determinar amb certesa qui és l'home i qui és la dona, segons les respostes que rep a les seves preguntes. Això seria fàcil si no fos perquè, en aquest joc, l'home menteix i fa veure que és la dona,

amb l'objectiu de confondre l'interrogador. Si, passats els cinc minuts, l'interrogador no és capaç de saber amb una certesa superior al 70 % qui és qui, llavors l'home guanya el joc, ja que ha aconseguit confondre l'interrogador fent-se passar per dona. Doncs bé, el test de Turing consisteix simplement a substituir, en aquest joc d'imitació, el paper de l'home per un ordinador. De tal manera que, si aconsegueix confondre l'interrogador, fent-li creure que és una persona, direm que l'ordinador és intel·ligent. Si bé és cert que fins ara no hi ha cap programa d'ordinador que hagi superat aquest test, cal dir que tampoc no és realment un objectiu dels investigadors en IA aconseguir superar-lo i, per tant, no s'hi han dedicat gaires esforços, exceptuant els diàlegs completament intranscendents del molt criticat premi Loebner.

El principal motiu de per què no és un objectiu en IA és que, segons l'estat actual de la IA, aquest joc d'imitació no és un bon indicador per determinar si una màquina és intel·ligent, ja que, com a màxim, només avalua aquells processos cognitius que són susceptibles de ser expressats verbalment. No obstant això, hi ha altres processos cognitius fonamentals que no són verbalitzables i la seva modelització i avaluació són imprescindibles en IA. L'exemple més paradigmàtic és l'actual recerca en robots autònoms amb l'objectiu de dotar-los de sofisticades habilitats sensorials i motores, que permetran que aquests robots puguin aprendre a reconèixer i comprendre el que vegin, toquin o escoltin. També hauran de tenir capacitats de raonament espacial per aprendre a interpretar el seu entorn, que generalment inclourà altres robots i també éssers humans, cosa que requerirà que també desenvolupin capacitats de socialització. Per poder mesurar els progressos cap a aquests objectius, un test com el proposat per Turing no serveix. Necessitem més aviat un conjunt de tests que avaluïn tot el rang de capacitats que conformen la intel·ligència, i en particular la capacitat d'adquirir coneixements de sentit comú, el problema més important que hem de resoldre per aconseguir intel·ligències artificials generals. En qualsevol cas, la repercussió més important del test de Turing és de caràcter filosòfic, ja que aquest test implica que, per decidir si una màquina és intel·ligent, l'important és observar externament si té un comportament intel·ligent o no, en lloc d'analitzar com les seves estructures i els seus mecanismes causals interns donen lloc a intel·ligència. Actualment aquesta és una qüestió que genera molta polèmica entre filòsofs de la ment, com per exemple l'existent entre John Searle i Paul i Patricia Churchland.

Els principals models en intel·ligència artificial: simbòlic, connexionista, evolutiu i corpori

Fins molt recentment, el model dominant en IA ha estat el simbòlic. Aquest model té les seves arrels en la hipòtesi SSF. De fet, encara segueix sent molt impor-

tant i actualment es considera el model «clàssic» en IA (també anomenat GOFAI, acrònim de Good Old Fashioned AI). És un model descendent (*top-down*) que es basa en el raonament lògic i la cerca heurística com a pilars per a la resolució de problemes, sense que el sistema intel·ligent necessiti formar part d'un cos ni estar situat en un entorn real. És a dir, la IA simbòlica opera amb representacions abstractes del món real que es modelen mitjançant llenguatges de representació basats principalment en la lògica matemàtica i les seves extensions. Per aquest motiu, els primers sistemes intel·ligents resolien principalment problemes que no requereixen interactuar directament amb l'entorn, com ara demostrar senzills teoremes matemàtics o jugar a escacs; els programes que juguen als escacs no necessiten, de fet, la percepció visual per veure les peces en el tauler ni actuadors per moure les peces. Això no vol dir que la IA simbòlica no pugui ser usada per programar el mòdul de raonament d'un robot físic situat en un entorn real, però en els primers anys de la IA no hi havia llenguatges de representació del coneixement ni de programació que permetessin fer-ho de manera eficient i per aquest motiu els primers sistemes intel·ligents es van limitar a resoldre problemes que no requerien interacció directa amb el món real. Actualment, la IA simbòlica es continua usant per demostrar teoremes o jugar a escacs, però també en aplicacions que requereixen percebre l'entorn i actuar-hi, com ara l'aprenentatge i la presa de decisions en robots autònoms.

Simultàniament, amb la IA simbòlica també va començar a desenvolupar-se una IA bioinspirada anomenada *connexionista*. Els sistemes connexionistes no són incompatibles amb la hipòtesi SSF, però, contràriament a la IA simbòlica, es tracta d'una modelització ascendent (*bottom-up*), ja que es basen en la hipòtesi que la intel·ligència emergeix a partir de l'activitat distribuïda d'un gran nombre d'unitats interconnectades que processen informació paral·lelament. A la IA connexionista, aquestes unitats són models molt aproximats de l'activitat elèctrica de les neurones biològiques.

L'any 1943, McCulloch i Pitts (McCulloch i Pitts, 1943) van proposar un model simplificat de neurona d'acord amb la idea que una neurona és essencialment una unitat lògica. Aquest model és una abstracció matemàtica amb entrades i sortides, que es correspondrien, respectivament, amb les dendrites i els axons. El valor de la sortida es calcula en funció del resultat d'una suma ponderada de les entrades, de manera que si aquesta suma supera un llindar preestablert llavors la sortida és un *1*, en cas contrari la sortida és *0*. Connectant la sortida de cada neurona amb les entrades d'altres neurones es forma una xarxa neuronal artificial. D'acord amb el que ja se sabia llavors sobre el reforçament de les sinapsis entre neurones biològiques, es va veure que aquestes xarxes neuronals artificials es podien entrenar per aprendre funcions que relacionessin les entrades amb les sortides mitjançant l'ajust dels pesos que serveixen per ponderar les connexions entre

neurones; per aquest motiu es va pensar que serien millors models per a l'aprenentatge, la cognició i la memòria, que els models basats en la IA simbòlica. No obstant això, els sistemes intel·ligents basats en el connexionisme tampoc no necessiten formar part d'un cos ni estar situats en un entorn real i, des d'aquest punt de vista, tenen les mateixes limitacions que els sistemes simbòlics. D'altra banda, les neurones reals posseeixen complexes arboritzacions dendrítiques amb propietats no tan sols elèctriques, sinó també químiques gens trivials. Poden contenir conductàncies iòniques que produeixen efectes no lineals i poden rebre desenes de milers de sinapsis que varien en posició, polaritat i magnitud. A més, el 90 % de les cèl·lules del cervell no són neurones, sinó les anomenades *cèl·lules glials*, que no solament regulen el funcionament de les neurones, ja que posseeixen potencials elèctrics, generen ones de calci i es comuniquen entre si, la qual cosa semblaria indicar que també tenen un paper molt important en els processos cognitius. No obstant això, no existeix cap model connexionista que inclogui aquestes cèl·lules; per tant, en el millor dels casos, aquests models són molt incomplets. En definitiva, tota l'enorme complexitat del cervell queda molt lluny dels models actuals i planteja dubtes sobre la utilitat de grans iniciatives com el projecte Human Brain de la UE. Aquesta immensa complexitat del cervell també fa pensar que l'anomenada singularitat, és a dir, futures superintel·ligències artificials que, basades en rèpliques del cervell, superessin la intel·ligència humana en un termini d'uns vint anys, és una predicció amb poc fonament científic.

Una altra modelització bioinspirada, també compatible amb la hipòtesi SSF, i no corpòria, és la computació evolutiva. Els èxits de la biologia en l'evolució d'organismes complexos va fer que al començament dels anys seixanta alguns investigadors es plantegessin la possibilitat d'imitar l'evolució per tal que els programes d'ordinador, mitjançant un procés evolutiu, milloressin automàticament les solucions als problemes per als quals havien estat programats. La idea és que aquests programes, gràcies a operadors de mutació i encreuament dels «cromosomes» que modelen els programes, produeixin noves generacions de programes modificats, de tal manera que les seves solucions siguin millors que les dels programes de les generacions anteriors. Atès que podem considerar que l'objectiu de la IA és la recerca de programes capaços de produir conductes intel·ligents, es va pensar que es podria fer servir la programació evolutiva per trobar aquests programes dins l'espai de programes possibles. La realitat és molt més complexa i aquesta aproximació té moltes limitacions, tot i que ha produït excel·lents resultats, en particular en la resolució de problemes d'optimització.

Una de les crítiques més fortes a aquests models no corporis es basa en el fet que un agent intel·ligent necessita un cos per poder tenir experiències directes amb el seu entorn (diríem que l'agent està «situat» en el seu entorn) en lloc que un programador proporcioni descripcions abstractes d'aquest entorn codificades

mitjançant un llenguatge de representació de coneixements. Sense un cos, aquestes representacions abstractes no tenen contingut semàntic per a la màquina. No obstant això, mitjançant la interacció directa amb l'entorn, l'agent pot relacionar els senyals que percep mitjançant els seus sensors amb representacions simbòliques generades a partir del que ha percebut. Alguns experts en IA, i en particular Rodney Brooks (Brooks, 1991), fins i tot van arribar a afirmar que no era ni tan sols necessari generar aquestes representacions internes, és a dir, que no cal que un agent hagi de tenir una representació interna del món que l'envolta, ja que el mateix món és el millor model possible de si mateix i la major part de les conductes intel·ligents no requereixen raonament, sinó que emergeixen a partir de la interacció entre l'agent i el seu entorn. Aquesta idea va generar molta polèmica i el mateix Brooks, uns quants anys més tard, va admetre que hi ha moltes situacions en les quals una representació interna del món és necessària perquè l'agent prengui decisions racionals.

L'any 1965, el filòsof Hubert Dreyfus va publicar un article titulat «Alchemy and artificial intelligence» (Dreyfus, 1965), on va afirmar que l'objectiu últim de la IA, és a dir la IA forta de tipus general, era tan inabastable com l'objectiu dels alquimistes del segle XVII, que pretenien transformar el plom en or. Dreyfus argumentava que el cervell processa la informació de manera global i contínua, mentre que un ordinador fa servir un conjunt finit i discret d'operacions deterministes, és a dir, aplicant regles a un conjunt finit de dades. En aquest aspecte podem veure un argument similar al de Searle, però Dreyfus, en posteriors articles i llibres (Dreyfus, 1992), va usar també un altre argument que defensa que el cos té un paper crucial en la intel·ligència. Va ser, doncs, un dels primers a advocar la necessitat que la intel·ligència formi part d'un cos amb el qual poder interaccionar amb el món. La idea principal és que la intel·ligència dels éssers vius deriva del fet d'estar situats en un entorn amb el qual poden interactuar gràcies als seus cossos. De fet, aquesta necessitat de corporeïtat està basada en la fenomenologia de Heidegger, que emfatitza la importància del cos amb les seves necessitats, desitjos, plaers, penes, forma de moure's, d'actuar, etc. Segons Dreyfus, la IA hauria de modelar tots aquests aspectes per assolir l'objectiu últim de la IA forta. És a dir, que Dreyfus no nega completament la possibilitat de la IA forta, però afirma que no és possible amb els mètodes clàssics de la IA simbòlica i no corpòria. Sens dubte es tracta d'una idea interessant que avui dia comparteixen molts investigadors en IA. Efectivament, l'aproximació corpòria amb representació interna ha anat guanyant terreny en la IA i actualment molts la considerem imprescindible per avançar cap a intel·ligències de tipus general. De fet, basem una gran part de la nostra intel·ligència en la nostra capacitat sensorial i motora. En altres paraules, el cos dona forma a la intel·ligència («the body shapes the way we think»), i, per tant, sense cos no hi pot haver intel·ligència de tipus general. Això és així perquè el

maquinari del cos, en particular els mecanismes del sistema sensorial i del sistema motor, determina el tipus d'interaccions que un agent pot realitzar. Al seu torn, aquestes interaccions conformen les habilitats cognitives dels agents i donen lloc a la cognició situada. És a dir, se situa la màquina en entorns reals, com passa amb els éssers humans, amb la finalitat que tingui experiències interactives que, eventualment, li permetin dur a terme alguna cosa similar al que proposa la teoria del desenvolupament cognitiu de Piaget (Inhelder i Piaget, 1958), segons la qual un ésser humà segueix un procés de maduració mental per etapes i potser els diferents passos d'aquest procés podrien servir de guia per dissenyar màquines intel·ligents. Aquestes idees han donat lloc a una nova subàrea de la IA anomenada *robòtica del desenvolupament* (Weng *et al.*, 2001).

De l'actual IA específica a una IA de tipus general?

Pràcticament tots els esforços en IA s'han centrat a construir intel·ligències artificials especialitzades i els èxits assolits en només seixanta anys d'existència són molt impressionants, sobretot durant l'últim decenni, principalment gràcies a la conjunció de dos elements: la disponibilitat d'enormes quantitats de dades i l'accés a la computació d'altes prestacions per poder analitzar-les. Efectivament, l'èxit de sistemes com ara AlphaGo (Silver *et al.*, 2016), Watson (Ferrucci *et al.*, 2013) i els avenços en vehicles autònoms han estat possibles gràcies a aquesta capacitat per analitzar grans quantitats de dades. No obstant això, pràcticament no hem avançat cap a la consecució de IA general. De fet, possiblement la lliçó més important que hem après al llarg dels seixanta anys d'existència de la IA és que el que semblava més difícil (diagnosticar malalties, jugar a escacs i a Go al més alt nivell) ha resultat ser relativament fàcil i el que semblava més fàcil ha resultat ser el més difícil. L'explicació a aquesta aparent contradicció cal buscar-la en la dificultat de dotar les màquines de coneixements de sentit comú.

Sense aquests coneixements no és possible una comprensió profunda del llenguatge ni una interpretació profunda del que capta un sistema de percepció visual, entre altres limitacions. De fet, el sentit comú és el requisit fonamental per aconseguir IA similar a la humana pel que fa a generalitat i profunditat. Els coneixements de sentit comú són fruit de vivències i experiències quan interactuem amb el nostre entorn, ja que la cognició humana és una cognició situada i corpòria. Les aproximacions no corpòries no permeten interaccions directes amb l'entorn, de manera que, inevitablement, donen lloc a problemes falsos i, per tant, a solucions falses. Tendeixen a definir els problemes en termes de tasques en un entorn especificades des d'una perspectiva abstracta d'objectes i relacions. Les capacitats cognitives no s'haurien d'estudiar fent abstracció del sistema sensorial i el sistema motor. Les capacitats més complicades d'assolir són les que requereixen interac-

cionar amb entorns no restringits ni prèviament preparats. Dissenyar sistemes que tinguin aquestes capacitats exigeix integrar desenvolupaments en moltes àrees de la IA. En particular, necessitem llenguatges de representació de coneixements que codifiquin informació sobre molts tipus diferents d'objectes, situacions, accions, etc., com també de les seves propietats i de les relacions entre ells. També necessitem nous algoritmes que, a partir d'aquestes representacions, puguin respondre, de manera robusta i eficient, preguntes sobre pràcticament qualsevol tema. Finalment, atès que necessitaran conèixer un nombre pràcticament il·limitat de coses, aquests sistemes han de ser capaços d'aprendre contínuament nous coneixements al llarg de tota la seva existència. En definitiva, és imprescindible dissenyar sistemes que integrin percepció, representació, raonament, acció i aprenentatge.

Els sistemes integrats: pas previ cap a la intel·ligència artificial general

En IA es parla de la metàfora de la catedral. Aquesta metàfora afirma que construir una IA de propòsit general és com construir una catedral. La construcció de la primera catedral va necessitar diverses generacions i per això la majoria dels que van treballar en la seva construcció no van arribar a veure-la acabada. Molts eren artesans que es van dedicar a construir maons cada vegada més perfectes i resistents que en el seu moment formarien part de la catedral. Actualment, i des de fa diverses dècades, els investigadors en IA estem construint també els «maons que compondran la catedral», és a dir, els algoritmes capaços de, per exemple, analitzar llenguatge, raonar, planificar i aprendre. I els millorem contínuament perquè facin cada cop més bé cadascuna d'aquestes funcions. Encara que la metàfora contingui elements vertaders, de fet, els humans, abans de poder construir una gran catedral, vam haver de construir molts altres edificis més senzills, com ara cabanes de fang i palla, posteriorment vam construir cases cada vegada més resistents i edificis cada cop més complexos, aprenent dels errors i fracassos al llarg de molts anys, fins que, finalment, vam aconseguir trobar els materials adequats per construir no tan sols grans catedrals, sinó també enormes gratacels, ponts i altres grans estructures. D'altra banda, és obvi que disposar de maons perfectes no és suficient per construir estructures complexes, també és imprescindible tenir una bona arquitectura. És a dir, que per molt sofisticats que siguin els algoritmes de raonament, planificació, aprenentatge, etcètera, mai no aconseguirem construir intel·ligències artificials generals si no sabem com integrar adequadament tots aquests elements. Només combinant aquests elements dins de sistemes cognitius integrats podrem començar a construir IA general, però la majoria de les investigacions actuals en IA continuen insistint a millorar els «maons» de manera massa aïllada, és a dir, sense col·laborar estretament amb aquells que tracten de construir els edificis, els científics que investiguen el que es coneix com *arquitectures cogni-*

tives. D'altra banda, si n'hi hagués més d'aquests últims, i menys constructors de maons, els progressos cap a la IA general serien més ràpids. Les arquitectures cognitives intenten resoldre el problema de com integrar els diferents components de la intel·ligència explorant hipòtesis sobre la naturalesa de la intel·ligència i les possibles interaccions entre aquests components. La primera proposta d'arquitectura cognitiva va ser el General Problem Solver (GPS) (Newell, Shaw i Simon, 1963). Molts anys després, Anderson i Lebiere (Anderson i Lebiere, 1998) van proposar l'arquitectura cognitiva ACT-R, inspirada en els treballs d'Allen Newell sobre una teoria unificada de la cognició, que, a més d'integrar teories d'atenció visual i moviment motor (per percebre i actuar en l'entorn), està dissenyada per poder modelar el procés de resolució de problemes en una varietat de tasques mitjançant una memòria declarativa, formada per peces de coneixement anomenades *chunks* (per exemple, «París és la capital de França»), i una altra memòria procedimental que conté coneixement sobre com actuar, expressat mitjançant un sistema de regles procedimentals del tipus «sí..., llavors...», similars a les del General Problem Solver. El sistema de producció selecciona aquella regla la condició de la qual es compleix en funció de l'estat del sistema i l'aplica. L'estat del sistema es modifica en temps real sobre la base dels mòduls d'atenció visual i moviment motor, com també del contingut de la memòria declarativa. L'ACT-R l'han utilitzat centenars d'investigadors, entre altres usos, per modelar com un ésser humà resol problemes com ara les torres de Hanoi, la resolució d'equacions algebraiques o la comprensió del llenguatge. Una altra arquitectura cognitiva clàssica, també fortament inspirada en el General Problem Solver (GPS), desenvolupada per John Laird al final dels vuitanta, és SOAR. En SOAR, la resolució d'un problema es basa en la recerca, dins l'espai de solucions possibles, d'un estat objectiu que representa una solució al problema. L'estratègia consisteix a aplicar les regles de producció per anar assolint subobjectius cap als quals moure's per tal d'apropar-se gradualment a l'estat solució. SOAR, a més, aprèn per experiència guardant les traces de les solucions que ha trobat resolent problemes per tal de poder-les reutilitzar en futurs problemes similars. Al llarg dels anys s'han desenvolupat una sèrie d'arquitectures SOAR, des de SOAR1 a mitjan anys vuitanta, fins a SOAR9 el 2008, cadascuna amb millores respecte de l'anterior. Sens dubte, SOAR i ACT-R són les arquitectures cognitives més importants, però no les úniques, posteriorment se n'han proposat d'altres com ara ICARUS (Langley i Choi, 2006), que, entre altres coses, també es basa en el concepte de sistema de producció, POLYSCHEME (Cassimatis, 2006), que, entre altres aspectes, fa servir el concepte *descomposició en subobjectius*, i la Companion Architecture (Forbus *et al.*, 2009), que es basa principalment en el raonament i l'aprenentatge per analogia. De totes, Companion és possiblement la més singular, ja que no pretén donar lloc a la construcció de sistemes intel·ligents completament autònoms, sinó, com el seu nom indica, a

sistemes d'ajuda que col·laborin amb els usuaris per resoldre problemes complexos, per exemple, recuperant de la seva memòria situacions precedents similars, ajudant a prendre decisions mostrant arguments a favor o en contra de possibles decisions alternatives, etc.

A més d'aquests desenvolupaments d'arquitectures cognitives, una bona part de les actuals investigacions en robòtica també són rellevants en tant que sistemes integrats, ja que inclouen components de raonament, planificació, aprenentatge, comunicació, percepció i acció. Els robots constitueixen, sens dubte, una plataforma excel·lent per a la investigació en sistemes integrats, un pas imprescindible cap a la IA de tipus general.

Intel·ligència artificial basada en l'anàlisi de dades massives

Entre les activitats futures, creiem que els temes de recerca més importants es continuaran centrant en el que, en anglès, es coneix com *massive data-driven AI*, és a dir, a explotar la possibilitat d'accedir a quantitats massives de dades i poder-les processar amb maquinari cada vegada més ràpid per tal de descobrir-hi relacions, detectar patrons i realitzar inferències i aprenentatge mitjançant models probabilístics com ara els sistemes d'aprenentatge profund (*deep learning*) (Bengio, 2009). No obstant això, aquests sistemes basats en l'anàlisi d'enormes quantitats de dades en el futur hauran d'incorporar mòduls que permetin explicar com s'ha arribat als resultats i a les conclusions que proposen, ja que la capacitat d'explicació és una característica irrenunciable en qualsevol sistema intel·ligent, perquè permet comprendre com funciona el sistema i avaluar la seva confiabilitat. Per altra banda, també és necessari per corregir possibles errors de programació i detectar si les dades d'entrenament estan esbiaixades. Cal saber si les respostes que ens donen són correctes per les raons correctes o a causa de coincidències que pot haver-hi en el conjunt de dades d'entrenament. Per això, un dels temes de recerca més importants en aprenentatge profund és dissenyar aproximacions interpretables d'aquests sistemes complexos d'aprenentatge. Una aproximació consisteix no tan sols a entrenar el sistema d'aprenentatge profund, sinó que, amb les mateixes dades, també s'entrena un altre sistema que el mimetitza usant una representació senzilla i transparent.

Un altre tema de recerca molt actual és la verificació i validació del programari que implementa l'algoritme d'aprenentatge. Això és especialment important en aplicacions d'alt risc com ara el pilotatge automàtic de vehicles autònoms. En aquests casos, necessitem una metodologia per provar i validar que aquests sistemes d'aprenentatge automàtic assoleixin alts nivells de precisió. Per exemple, podríem exigir garanties sobre el reconeixement d'obstacles per tal que la probabilitat de col·lisió sigui inferior a una col·lisió per cada 10 milions de quilòmetres.

Actualment s'estan explorant algunes idees prometedores, com per exemple l'ús de simuladors per generar proves sistemàtiques, però cal tenir en compte la gran dificultat de construir simuladors realistes.

Una altra idea es coneix com *aprenentatge adversari* (*adversarial learning*) i consisteix a entrenar un segon sistema de IA que tracta de «trencar» el nostre programari d'aprenentatge intentant trobar-ne els punts febles. Per exemple, en el cas del reconeixement visual, generant imatges que provoquin que el sistema prengui la decisió equivocada. Hi ha també altres aproximacions al problema de la verificació i validació basades en tècniques més clàssiques de l'enginyeria del programari.

Altres temes clau en IA

Altres àrees de la IA que continuaran sent objecte d'investigació extensiva són els sistemes multiagent, la planificació d'accions, el raonament basat en l'experiència, la visió artificial, la comunicació multimodal persona-màquina, la robòtica humanoide, la robòtica social i les noves tendències en robòtica del desenvolupament que poden ser clau per dotar les màquines de sentit comú. També veurem progressos significatius gràcies a les aproximacions biomimètiques per reproduir en màquines el comportament d'animals. No es tracta únicament de reproduir el comportament d'un animal, sinó de comprendre com funciona el cervell que produeix aquest comportament. Es tracta de construir i programar circuits electrònics que reprodueixin l'activitat cerebral que genera aquest comportament. Alguns biòlegs estan interessats en els intents de fabricar un cervell artificial tan complex com sigui possible, perquè consideren que és una manera de comprendre millor l'òrgan i els enginyers busquen informació biològica per fer dissenys més eficaços. Mitjançant la biologia molecular i els avenços recents en optogenètica, serà possible identificar quins gens i quines neurones tenen un paper clau en les diferents activitats cognitives.

Una altra àrea d'interès important per a la IA, i en particular per a la robòtica, és la ciència de materials. Per exemple, per al desenvolupament de músculs artificials, una possible tecnologia consisteix a intercalar capes de cautxú de silici amb capes de polímer electroactiu, de tal manera que el conjunt es flexioni en aplicar un camp elèctric (Anderson *et al.*, 2014). També hi ha resultats interessants en cartílags artificials construïts mitjançant filaments de polímers amb molècules que atrauen l'aigua per mimetitzar les propietats dels cartílags naturals (Chen *et al.*, 2009). Finalment, s'estan usant compostos de cautxú de silici carregat amb nanopartícules de níquel per a pells artificials, ja que la resistència del compost disminueix amb la pressió a què és sotmès, o també sensors capacitius, és a dir, la capacitat elèctrica canvia amb la pressió (Schmitz *et al.*, 2011).

Aquesta aproximació pluridisciplinària a la IA inspirada en la biologia i l'ús de resultats en ciència de materials pot produir un efecte sinèrgic que canviï profundament la naturalesa de la IA i fins i tot potser la nostra comprensió del que és la intel·ligència.

Pel que fa a les aplicacions, algunes de les més importants continuaran sent les relacionades amb el *web*, els videojocs i els robots autònoms (en particular vehicles autònoms, robots socials, robots per a l'exploració de planetes, etc.). Les aplicacions en el medi ambient i l'estalvi energètic també seran importants, com també en l'economia i la sociologia.

Finalment, les aplicacions de la IA en l'art (arts visuals, música, dansa, narrativa) canviaran de manera important la naturalesa del procés creatiu. Els ordinadors ja no són només eines d'ajuda a la creació, els ordinadors comencen a ser agents creatius. Això ha donat lloc a una nova i molt prometedora àrea d'aplicació de la intel·ligència artificial anomenada *creativitat computacional*, que ja ha produït resultats molt interessants (Colton *et al.*, 2009, 2015) (López de Mántaras, 2016) en música, arts plàstiques i narrativa, entre altres activitats creatives. Però possiblement l'aspecte més interessant de la recerca en creativitat computacional és el seu potencial d'augmentar la creativitat humana i fins i tot de posar-la a l'abast de gairebé tothom, és a dir, quelcom que podríem anomenar *democratització de la creativitat*. Efectivament, hi ha una nova tendència coneguda amb el nom de *creació assistida*, que té implicacions importants per a la creativitat: d'una banda, els sistemes de creació assistida fan que un ampli ventall d'habilitats creatives siguin més accessibles. De l'altra, plataformes col·laboratives com la que hem desenvolupat dins del projecte europeu PRAISE per aprendre música (Yee-King i Inverno, 2014) faciliten l'aprenentatge de noves habilitats creatives. PRAISE és una plataforma d'aprenentatge basada en la xarxa social que inclou humans i agents artificials que donen retroacció (*feedback*) a un estudiant de música pel que fa a composició, arranjament i interpretació musical. Els alumnes pugen les seves solucions a la plataforma seguint el pla de lliçons proposat per un tutor (composicions, arranjaments o interpretacions). A continuació, els agents intel·ligents i altres estudiants i tutors analitzen aquestes solucions i en fan comentaris. Per exemple, en el cas d'una composició, l'agent podria dir: «la vostra modulació és molt bona, però podeu intentar modular-ho tot una tercera major cap amunt en els compassos 5 a 8».

En el cas de les interpretacions, altres agents de programari intel·ligent comparen les de l'alumne amb una de prèviament gravada pel tutor. Una càmera captura el gest de l'estudiant i els agents intel·ligents també fan comentaris sobre possibles postures incorrectes. Eines com aquesta que acceleren el temps d'adquisició d'habilitats donen lloc a un fenomen de *democratització de la creativitat*.

L'any 1962, Douglas Engelbart (Engelbart, 1962) ja va escriure sobre una «màquina d'escriptura que permetria utilitzar un nou procés de composició de textos». Engelbart argumentava que «es podrien integrar les idees més fàcilment i, per tant, aprofitar la creativitat humana de manera contínua». La visió d'Engelbart no tan sols tractava d'augmentar la creativitat individual. També volia augmentar la intel·ligència col·lectiva i la creativitat dels grups a través de la millora de la col·laboració i la capacitat de resolució de problemes del grup. Una idea bàsica és que la creativitat és un procés social que es pot augmentar a través de la tecnologia. En projectar aquestes idees cap al futur, podríem imaginar un món on la creativitat és altament accessible i (gairebé) qualsevol pot esdevenir un molt bon escriptor, o pintar com els grans mestres, o compondre música d'alta qualitat i fins i tot descobrir noves formes d'expressió creativa. Encara que això és actualment pura ficció, ja hi ha alguns exemples de creació assistida. Un dels més interessants és el sistema de percussió assistit desenvolupat a l'Institut de Tecnologia de Geòrgia (Bretan i Weinberg, 2016). Consisteix en una extremitat robòtica que permet tocar la bateria amb tres braços. El braç robòtic es pot connectar a l'espatlla del músic i respon als seus gestos i a la música que sent.

Un altre resultat molt interessant en la creativitat assistida és l'anàlisi de l'estil musical i la transferència d'harmonies entre gèneres musicals. El sistema, desenvolupat al SONY Computer Science Laboratories de París (Martín *et al.*, 2015; Papadopoulos *et al.*, 2016), ajuda els compositors a harmonitzar una obra musical corresponent a un gènere segons l'estil d'un altre gènere completament diferent. Per exemple, permet harmonitzar un estàndard de *jazz* segons l'estil de Mozart.

Algunes conclusions

En aquest article hem començat explicant la hipòtesi dels sistemes de símbols físics i la distinció entre la IA feble i específica i la IA forta i general, deixant clar que totes les intel·ligències artificials actuals són del tipus feble i específic. A continuació hem parlat dels escrits d'Alan Turing sobre la IA, emfatitzant el caràcter visionari de les seves idees. Després hem descrit els quatre models principals de la IA, és a dir, simbòlic, connexionista, evolutiu i corpori, i hem argumentat la importància cabdal del model corpori per assolir veritables intel·ligències artificials. Després hem parlat de la importància de dotar de coneixements de sentit comú les màquines, per tal de passar de les intel·ligències específiques a les generals. També hem remarcat la importància i la necessitat de dissenyar sistemes integrats com a pas previ cap a intel·ligències de tipus general. Finalment hem parlat de les noves tendències en IA basades en l'anàlisi de dades massives i hem acabat amb altres temes clau en IA com ara la creativitat computacional i la relació de la IA amb altres ciències.

Reflexió final

Per molt intel·ligents que arribin a ser les futures intel·ligències artificials, en particular les de tipus general, mai no seran iguals que les intel·ligències humanes, ja que, com hem argumentat, el desenvolupament mental que requereix tota intel·ligència complexa depèn de les interaccions amb l'entorn i aquestes interaccions depenen al seu torn del cos, en particular del sistema perceptiu i del sistema motor. Això, juntament amb el fet que les màquines no seguiran processos de socialització i culturització com els nostres, incideix encara més en la qüestió que, per molt sofisticades que arribin a ser, seran intel·ligències diferents de les nostres. El fet de ser intel·ligències alienes a la humana i, per tant, alienes als valors i a les necessitats humanes ens hauria de fer reflexionar sobre possibles limitacions ètiques al desenvolupament de la intel·ligència artificial. En particular, estem d'acord amb Weizenbaum (Weizenbaum, 1976) que cap màquina no hauria de prendre decisions de manera completament autònoma o donar consells que requereixin, entre altres coses, saviesa, producte d'experiències humanes, com també tenir en compte valors humans. Exemples clars d'usos de la IA que no s'haurien de permetre per motius ètics són les armes autònomes i els sistemes informàtics que operen en borsa prenent decisions de compra i venda en fraccions de segon i les aplicacions que atempten contra la nostra privacitat. Actualment els algoritmes en què es basen els motors de cerca a Internet, els sistemes de recomanació i els assistents personals dels nostres telèfons mòbils, coneixen prou bé el que fem, les nostres preferències i els nostres gustos i fins i tot poden arribar a inferir el que pensem i com ens sentim. L'accés a quantitats massives d'informació, que voluntàriament generem, és fonamental perquè això sigui possible, ja que mitjançant l'anàlisi d'aquestes dades provinents de fonts diverses és possible trobar relacions i patrons que serien impossibles de detectar sense les tècniques de IA. Tot això provoca una pèrdua alarmant de privacitat. Per evitar-ho hauríem de tenir dret a posseir una còpia de totes les dades personals que generem, controlar-ne l'ús i decidir a qui permetem l'accés i sota quines condicions, en lloc que estiguin en mans de grans corporacions sense poder saber quin ús en fan.

La IA està basada en programació complexa, i, per tant, necessàriament cometrà errors. Però fins i tot suposant que fos possible desenvolupar programari completament fiable, hi ha dilemes ètics que els desenvolupadors de programari han de tenir en compte a l'hora de dissenyar-lo. Per exemple, un vehicle autònom podria decidir atropellar un vianant per evitar una col·lisió que podria causar danys als seus ocupants. Equipar les empreses amb sistemes avançats de IA per fer la gestió i la producció més eficients requerirà menys empleats humans i, per tant, generarà més atur. Aquests dilemes ètics fan que molts experts en IA assenyalin la necessitat de regular-ne el desenvolupament. En alguns casos s'hauria fins i tot de

prohibir l'ús de la IA. Un exemple clar són les armes autònomes. Els tres principis bàsics que regeixen els conflictes armats: *discriminació* (la necessitat de discernir entre combatents i civils o entre un combatent rendint-se i un disposat a atacar), *proporcionalitat* (fins a quin punt són acceptables els danys col·laterals) i *precaució* (minimització del nombre de víctimes) són extraordinàriament difícils d'avaluar, i, per tant, gairebé impossibles de complir pels sistemes de IA que controlen les armes autònomes. Fins i tot en el cas que a molt llarg termini les màquines tinguessin aquestes capacitats, seria indigne delegar en una màquina la decisió de matar. Però, a més de regular, és imprescindible educar els ciutadans sobre els riscos de les tecnologies intel·ligents i dotar-los de les competències necessàries per controlar-les en lloc de ser ells els controlats. Necessitem futurs ciutadans molt més informats, amb més capacitat per avaluar els riscos tecnològics, amb més sentit crític i disposats a fer valer els seus drets. Aquest procés de formació ha de començar a l'escola i tenir continuació a la universitat. En particular, cal que els estudiants de ciència i enginyeria rebin una formació ètica que els permeti comprendre millor les implicacions socials de les tecnologies que molt probablement desenvoluparan. Només si invertim en educació aconseguirem una societat que pugui aprofitar els avantatges de les tecnologies intel·ligents i minimitzar-ne els riscos.

El camí cap a la IA de tipus general seguirà sent llarg i difícil, al cap i a la fi la IA té només seixanta anys d'existència i, com diria Carl Sagan, seixanta anys són un brevíssim moment en l'escala còsmica del temps; o, com molt poèticament va dir Gabriel García Márquez: «Des de l'aparició de vida visible a la Terra van haver de transcórrer 380 milions d'anys perquè una papallona aprengués a volar, 180 milions d'anys més per fabricar una rosa sense altre compromís que el de ser formosa, i quatre eres geològiques perquè els éssers humans fossin capaços de cantar millor que els ocells i morir-se d'amor».

Agraïments

Vull retre un reconeixement als meus mestres el doctor Josep Aguilar-Martin, el professor Enric Trillas i el professor Lotfi Zadeh pels seus ensenyaments; sense ells segurament no hauria tingut l'oportunitat de començar el meu fascinant viatge en el món de la recerca i en particular en la intel·ligència artificial. També vull agrair a tots els meus exdoctorands, companys de l'Institut d'Investigació en Intel·ligència Artificial (IIIA) i col·legues d'arreu del món la seva dedicació i col·laboració, ja que sense ells no hauria pogut continuar aquest viatge que ja fa quaranta anys que dura. Finalment, el meu molt sentit agraïment a la meva família i en particular a la meva esposa (q. e. p. d.) i als meus fills per la seva infinita paciència quan el meu viatge em duia lluny de casa.

Referències bibliogràfiques

Anderson, Ch.; Lebiere, J. R. (1998). *The atomic components of thought*. Nova Jersey: Erlbaum Pub.

Anderson, I.; Rosset, S.; McKay, T.; Shea, H. (2014). «Stack design for portable artificial muscle generators: is it dangerous to be short and fat?». *Proceedings of SPIE* [San Diego], vol. 9056: *Electroactive polymer actuators and devices* (EAPAD).

Bengio, Y. (2009). «Learning deep architectures for AI». *Foundations and Trends in Machine Learning*, vol. 2 (1), p. 1-127.

Bretan, M.; Weinberg, G. (2016). «A survey of robotic musicianship». *Commun. ACM*, vol. 59 (5), p. 100-109.

Brooks, R. A. (1991). «Intelligence without reason». A: *IJCAI'91 Proceedings of the 12th International Joint Conference on Artificial Intelligence*. Vol. 1. Sidney: AAAI Press, p. 569-595.

Cassimatis, N. (2006). «A cognitive substrate for human-level intelligence». *AI Magazine*, vol. 27, p. 45-56.

Chen, M.; Briscoe, W. H.; Armes, S. P.; Klein, J. (2009). «Lubrication at physiological pressures by polyzwitterionic brushes». *Science*, vol. 323 (5922), p. 1698-1701.

Colton, S.; Halskov, J.; Ventura, D.; Gouldstone, I.; Cook, M.; Pérez-Ferrer, B. (2015). «The painting fool sees! New projects with the automated painter». A: *Proceedings of the Sixth International Conference on Computational Creativity (ICCC 2015)*. Provo, Utah: Brigham Young University, p. 189-196.

Colton, S.; López de Mántaras, R.; Stock, O. (2009). «Computational creativity: coming of age». *AI Magazine*, vol. 30 (3), p. 11-14.

Dreyfus, H. L. (1965). «Alchemy and artificial intelligence». *RAND Paper* [Califòrnia: The RAND Corporation], P-3244.

— (1992). *What computers still can't do: a critique of artificial reason*. Cambridge, MA: MIT Press.

Engelbart, D. C. (1962). «Augmenting human intellect: a conceptual framework». *Technical Report* [Stanford: Stanford Research Institute] (octubre), p. 5.

Ferrucci, D. A.; Levas, A.; Bagchi, S.; Gondek, D.; Mueller, E. T. (2013). «Watson: Beyond Jeopardy!». *Artif. Intell.*, vol. 199, p. 93-105.

Forbus, K.; Klenk, M.; Hinrichs, T. (2009). «Companion cognitive systems: design goals and lessons learned so far». *IEEE Intelligent Systems*, vol. 24, p. 36-46.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. Nova York: John Wiley.

Inhelder, B.; Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence.* Nova York: Basic Books.

Langley, P.; Choi, D. (2006). «A unified cognitive architecture for physical agents». A: *Proceedings of the Twenty-First AAAI Conference on Artificial Intelligence.* Pittsburgh: AAAI Press.

López de Mántaras, R. (2016). «Artificial intelligence and the arts: toward computational creativity». A: *The next step*: *Exponential life*. Madrid: BBVA, p. 100-125. (BBVA Open Mind)

Martín, D.; Frantz, B.; Pachet, F. (2015). «Improving music composition through peer feedback: experiment and preliminary results». A: Steels, L. (ed.). *Music learning with massive open online courses (MOOCs)*. Amsterdam: IOS Press, p. 195-204.

McCulloch, W. S.; Pitts, W. (1943). «A logical calculus of ideas immanent in nervous activity». *Bulletin of Mathematical Biophysics*, vol. 5, p. 115-133.

Newell, A.; Shaw, J. C.; Simon, H. A. (1963). «Report on a general problem-solving program». A: *Proceedings of the International Conference on Information Processing*. París: UNESCO House, p. 256-264. També publicat a: Feigenbaum, E. A.; Feldman, J. (ed.). *Computers and thought*. Nova York: McGraw Hill, p. 279-293.

Newell, A.; Simon, H. (1976). «Computer science as empirical inquiry: symbols and search». *Communications of the ACM*, vol. 19 (3), p. 113-126.

Papadopoulos, A.; Roy, P.; Pachet, F. (2016). «Assisted lead sheet composition using FlowComposer». A: *Principles and practice of constraint programming. Proceedings of the 22nd International Conference, CP 2016* (Tolosa, França). Berlín: Springer (Lecture Notes in Computer Science. Springer Book Series, núm. 9892).

Schmitz, A.; Maiolino, P.; Maggiali, M.; Natale, L.; Cannata, G.; Metta, G. (2011). «Methods and technologies for the implementation of large-scale robot tactile sensors». *IEEE Transactions on Robotics*, vol. 27 (3), p. 389-400.

Searle, J. R. (1980). «Minds, brains, and programs». *Behavioral and Brain Sciences*, vol. 3 (3), p. 417-457.

Silver, D.; Huang, A.; Maddison, C. J.; Guez, A.; Sifre, L.; Driessche, G. van den; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; Dieleman, S.; Grewe, D.; Nham, J.; Kalchbrenner, N.; Sutskever, I.; Lillicrap, T.; Leach, M.; Kavukcuoglu, K.; Graepel, T.; Hassabis, D. (2016). «Mastering the game of go with deep neural networks and tree search». *Nature*, vol. 529 (7587), p. 484-489.

Turing, A. M. (1948). «Intelligent machinery». *National Physical Laboratory Report*. També publicat a: Meltzer, B.; Michie, D. (ed.) (1969). *Machine intelligence 5*. Edimburg: Edinburgh University Press, p. 3-23.

— (1950). «Computing machinery and intelligence». *Mind*, vol. lix (236), p. 433-460.

Weizenbaum, J. (1976). *Computer power and human reason*: *From judgment to calculation*. San Francisco: W. H. Freeman and Co.

Weng, J.; McClelland, J.; Pentland, A.; Sporns, O.; Stockman, I.; Sur, M.; Thelen, E. (2001). «Autonomous mental development by robots and animals». *Science*, vol. 291, p. 599-600.

Yee-King, M.; Inverno, M. d' (2014). «Pedagogical agents for social music learning in Crowd-based Socio-Cognitive Systems». A: *Proceedings First International Workshop on the Multiagent Foundations of Social Computing, AAMAS-2014*. París: IFAAMAS.


INSTITVT D'ESTVDIS CATALANS
MCMVII


9 788499 654003