

L'Anti-Turing

Enric Plaza



La làpida Atua-Mata-Riri de Rapa Nui (illa de Pasqua) mostra l'escriptura bustrofèdica anomenada *rongo-rongo*, encara sense desxifrar.

Fa cinquanta-sis anys

Fa cinquanta-sis anys que A.M. Turing va escriure el dignament famós article *Computing Machinery and Intelligence*, (<http://www.loebner.net/Prize/TuringArticle.html>). Publicat el 1950 a la revista *Mind*, volum 59, p. 433-460 s'avança així sis anys a la instauració del programa de recerca de la Intel·ligència Artificial a Dartmouth el 1956. L'article proposa el “test de Turing”, possiblement la idea de la IA més coneguda per la societat —si més no a la societat anglosaxona, no m'atreveixo a afirmar-ho del regne del *qu'inventen ellos*.

L'article d'en Turing no només presenta la idea del test, presenta les qüestions que hom discuteix sobre la possibilitat de tenir intel·ligència en artefactes artificials. En aquest sentit, l'article és fascinant, car adreça totes les crítiques habituals a la possibilitat de tenir intel·ligència artificial. Així, Turing revisa i contesta la objecció teològica (ni animals ni màquines tenen ànima), la matemàtica (els límits a la calculabilitat del teorema de Gödel), la de la consciència (la manca de consciència i emocions), l'argument de Lady Lovelace (les màquines només poden fer el que els diem), i els diferent arguments del tipus “les màquines mai no podran fer X” (on X pot ser riure un acudit, aprendre de l'experiència, o gaudir d'unes maduixes amb crema).

Llegint l'article observem que els punts de desacord en la discussió no han canviat gens ni gaire en cinquanta-sis anys. Per exemple, el “gran debat” ocasionat pel llibre de Roger Penrose “The Emperor's New Mind” es deriva de la raó esgrimida per atacar la IA és la limitació al poder deductiu derivada del teorema de Gödel —és a dir la objecció matemàtica ja revisada per Turing. En escrits posteriors, Penrose discuteix diverses possibilitats que podrien superar les seves objeccions; una de les més interessants és quan Penrose planteja la capacitat d'aprenentatge a partir de la interacció amb l'entorn. Precisament la capacitat d'aprenentatge és l'argument que Turing usa contra l'argument de Lady Lovelace (les màquines només poden fer el que els diem).

Sens dubte, Turing fa un article seminal que planteja, des d'un punt a la història on tot el referent a les màquines digitals és encara a les beceroles, les qüestions fonamentals que aparentment encara arrosseguem com problemes irresolts. Tanmateix, la discussió de Turing d'aquestes qüestions es basa en la seva proposta del "joc d'imitació" (el que habitualment s'anomena Test de Turing). El joc d'imitació presenta un test d'indistingibilitat entre (A) un ésser humà i (B) una màquina que dialoguen (possiblement per escrit, en el que avui en diem un xat) amb (C) un interrogador. Si C no pot distingir d'una manera estadísticament significativa A de B direm que són indistingibles; la conseqüència segons Turing és que podrem respondre de manera positiva la pregunta "Poden pensar les màquines?" (que és equivalent a "poden ser intel·ligents les màquines").

El test a prova

Donat que els problemes que Turing ja provava de contestar (i resoldre) encara són vius i irresolts avui dia, potser el test proposat per Turing no ha estat tan útil com ell hauria volgut. Per tant, potser hauríem de definir un altre test —o potser hauríem de canviar la pregunta. Així, en lloc de preguntar "Poden pensar les màquines?" (que és equivalent a "poden ser intel·ligents les màquines d'una manera igual als humans") podríem definir una nova pregunta més general, per exemple "Com podem reconèixer la intel·ligència amb qualsevol objecte present a l'univers?". Abans de redefinir el test cal analitzar primer el joc d'imitació proposat per Turing i trobar els problemes que en són inherents.

Un primer problema és que es basa en el xauvinisme humà i crida a l'engany. M'explico: sabent les regles del joc és molt fàcil detectar si la màquina és A o B tot preguntant-li sobre la seva vida íntima. En aquesta circumstància és molt fàcil detectar qui és la màquina a menys que menteixi (per exemple, digui que té una muller i dues filles). La manera d'evitar això seria restringir els temes de conversa, però aquesta limitació restaria vàlida al test. El problema aquí és el xauvinisme humà: per tal com només admetem com intel·ligència aquella que és humana (o prou semblant a la humana), la màquina que se sotmet al test tindrà com objectiu "imitar" els humans més que mostrar-se tal com és.

L'altre problema és que el test de Turing és booleà, és a dir que pretén respondre sí o no a la qüestió de la intel·ligència: la tens si sembles humà i no altrament. El mateix Turing, en la seva resposta a l'objecció teològica (ni animals ni màquines tenen ànima) troba insatisfactori separar humans i animals: tots venim de la mateixa nissaga i seria més adequat parlar dels diferents tipus d'intel·ligència que tenen els animals. Aquesta intuïció de Turing és molt encertada, i hauria d'incloure's en qualsevol nou test que es proposés, però caldria una resposta no booleana (i per tant diferent de la del test de Turing). Seguir per aquest camí requereix desenvolupar una taxonomia dels tipus d'intel·ligència, cosa que s'ha demostrat problemàtica en la biologia actual. Així, per bé que s'accepta que hi ha "comunicació" animal, es discuteix si els animals "enten" els signes o senyals en els quals es basa la comunicació o si simplement "responen" amb la conducta adequada.

D'animals i d'aliens

De fet, aquesta qüestió és fortament relacionada amb el test de Turing: la hipòtesi subjacent és que si pot comunicar-se com un humà necessàriament tindrà la capacitat de comprensió que

anomenem intel·ligència (o “enteniment”). Una novetat important el 2006 és la descoberta (o potser la *constatació*) que la comunicació animal posseeix *sintaxi*. En efecte, hi ha proves que la capacitat d’enunciar “frases” no és exclusiva dels humans, ans compartida amb els simis. En concret, s’ha observat que els micos de nas espàtula (*Cercopithecus Nictitans*) de Nigèria prenen elements discrets de comunicació (mots) i construeixen amb ells una frase on el significat global és diferent del dels “mots” individuals. Per exemple, els micos de nas espàtula tenen un “so d’alerta” anomenat *pyow* que avisa de la presència d’un predador específic (el lleopard) i un so *hack* que s’usa quan hi vola aprop una àliga. Tanmateix, s’ha comprovat que usen una frase composta de tres *pyow* seguits de fins a quatre *hacks* que simplement vol dir “Anem-hi”; l’elocució d’aquesta frase no té res a veure amb lleopards ni àligues, és una ordre que fa que tot el grup es posi en moviment. També se sap que els dofins usen noms propis pels individus. Malgrat les resistències, s’obre pas l’estudi seriós de la cognició animal — i segur que el debat social encetat per la IA sobre la natura del que cal considerar intel·ligència hi ha tingut molt a veure.



Una problemàtica no idèntica però molt relacionada ens pot ajudar en aquest tràngol: la qüestió de la (comunicació amb) vida intel·ligent extraterrestre. Enteneu-me, no es tracta de la qüestió de si hi ha vida (o vida intel·ligent) a altres planetes, sinó més aviat de quin criteri usariem per determinar quins éssers són “intel·ligents” o “pensen”. La solució a primer cop d’ull sembla senzilla, es tractaria d’observar si tenen instruments, ciutats o naus espacials — usar un criteri social en lloc d’individual.

Però aquest criteri és defectuós, car intent determinar si s'assemblen prou a nosaltres o no. Aquest criteri no detectaria altres formes de civilització basades, per exemple, en “intel·ligència d'eixam”; aquest xoc cultural es dona a la pel·lícula *Starship Troopers* (Paul Verhoeven 1997, basada en una novel·la d'en Robert Heinlein) on els humans ocupen un planeta i ataquen els seus habitants de tipus “colònies d'insectes” perquè no reconeixen que es tracte d'una “civilització”. Naturalment, quan la civilització d'insectes contraataca la percepció canvia. Ja és massa tard, però ens dona el primer criteri, insuficient però clar, per detectar vida intel·ligent: la guerra.



Aquest problema ja es va donar històricament un cop: durant el descobriment del Nou Món (ara Amèrica) es va debatre, amb conseqüències pràctiques molt dramàtiques, si els éssers d'allà eren “humans” (en els termes de la època es discutia si tenien ànima i tenien per tant dret a ser tractats com homes o no). Pot semblar que els europeus d'aleshores eren simplement estúpids i bàrbars, però això menystindria el xoc cultural que varen patir en arribar al Nou Món (la millor pel·lícula del 2006 és *The New World* i retrata a la perfecció l'astorament mutu d'indígenes i europeus).

Recordem que els europeus havien considerat africans, hindis i xinesos humans (cosa que no evitava llur explotació, és clar), però en trobar uns éssers tan diferents, en un lloc tan estrany, varen dubtar. La raó és, com abans, l'ús d'un criteri de similitud respecte “nosaltres”.

Per tant, si deixem a part la guerra, hauríem de intentar definir un criteri de reconeixement de la intel·ligència aliena basat en la comunicació. Per tal de fer això, cal considerar altre cop què vol dir *comunicar-se*, què vol dir *entendre* els signes que s'usen en la comunicació (cosa que sembla crear un raonament circular: definir “intel·ligència” a partir de la capacitat d'entendre!). Per això, faré servir un experiment mental proposat per Douglas Hofstadter al llibre *GEB*, també conegut com *Gödel, Escher, Bach*; tanmateix, el meu reconté és prou diferent del seu i no s'hauria d'entendre que els meus possibles errors són culpa del GEB.

Doncs bé, l'experiment mental contempla la qüestió de com una espècie extraterrestre podria entendre (o no) una certa informació codificada; en concret, ell proposa que trobin un disc que enregistra una sonata de Bach. Què voldria dir que uns aliens “entenen” el disc de Bach? En primer lloc, llur fisiologia pot ser molt diferent, i el seu sistema auditiu molt allunyat del nostre o inexistent. En segon lloc, el registre (digital o analògic) no està clar que “representi” alguna “senyal” de tipus sonor. En tercer lloc, les emocions o efectes que la música ens provoca a nosaltres no han de ser compartides amb els aliens. Donats aquests problemes tan grans, és doncs impossible comunicar-se o (si més no) establir la possibilitat de reconèixer l'altre malgrat els problemes de comunicació?

Hofstadter proposa tres nivells de comprensió d'un missatge basant-se en el concepte de descodificació. Els tres nivells són: 1) el marc del missatge, 2) el missatge extern, i 3) el missatge intern. El marc del missatge és aquell fa reconèixer un objecte com “portador d'informació”, és a dir reconèixer que es pot establir un procés de descodificació sobre aquell objecte. Els nostres aliens podrien resoldre el problema del marc si troben en disc dins un satèl·lit, reconeixen que disc i satèl·lit són artefactes construïts per algú, i fan la hipòtesi que el disc pot “contenir” alguna mena de “missatge”. Aquest reconeixement crea un context totalment nou quant a la percepció de què és el disc. Examinant el seu “contingut”, encara que no entenguin la codificació de la música, poden adonar-se que conté un senyal, és a dir que no té un contingut aleatori, sinó que es donen certes regularitats. En conseqüència, saben que el disc conté un codi, per bé que no sàpiguen descodificar aquest codi.

Un cas semblant és l'exemple d'escriptura *rongo-rongo* de Rapa Nui mostrat a la figura 1. L'escriptura rongo-rongo no s'ha desxifrat mai, però tot i així la reconeixem com a “codi”. Per tant, hem resolt només el nivell 1 dels tres nivells de comprensió del missatge. El nivell 2 s'assoleix si tenim una manera de comprendre la sintaxi (per exemple, trobar la Pedra Rossetta del *rongo-rongo*; això ens permetria construir un procés de descodificació). El nivell tres, el missatge intern, és el contingut que es desitjava transmetre: el significat d'un text o les experiències sensorials ocasionades per la sonata de Bach.

Sorprenentment, aquest petit esquema de tres nivells ens permet resoldre el nostre problema de com reconèixer intel·ligències alienes, siguin animals o extraterrestres: es tracta simplement de resoldre el problema del marc del missatge. En efecte, els aliens potser no “entendran totalment” la sonata de Bach, però poden entendre que és alguna mena de missatge perquè s'han adonat que hi ha un contingut *codificat*. Reconèixer el *codi* és reconèixer la intel·ligència de l'altre que és

diferent de nosaltres. Descobrir que els micos de nas espàtula usen sintaxi ens permet veure que tenen un tipus d'intel·ligència pròpia — diferent de la nostra però tanmateix real.

Enric Plaza

enric@iia.csic.es

PS: El codi genètic *també* és un *codi*, però això ho deixarem per un altre Cau del Hàcker.