

**Pasado, presente e futuro  
da intelixencia artificial**

**Conversa con Marvin Minsky**



**N**o mes de outubro de hai dous anos o profesor Ramón López de Mántaras foi invitado polo Media Lab do Massachusetts Institute of Technology (MIT) a participar nun simposio para celebrar o artigo “Music, Mind, and Meaning”, do profesor Marvin Minsky.

Un día despois foi convidado a cear na casa da familia Minsky, o que lle deu oportunidade para preguntarlle pola súa vida e as súas contribucións, así como polo pasado, presente e futuro da intelixencia artificial (IA).

Marvin Minsky fixo transcendentais contribucións en campos como a IA, as ciencias cognitivas, as matemáticas e a robótica, amén de pioneiro na análise das redes neuronais artificiais.

Esta entrevista, publicada orixinariamente no *Butlletí de l'Associació Catalana d'Intel·ligència Artificial*, é o resultado das notas que tomou, ademais do que gardou na memoria da súa conversa.

Durante máis ou menos as tres últimas décadas houbo pouco emprego interesante en IA debido a esta mudanza de obxectivos cara á construción de sistemas de IA demasiado específicos. Temos unha morea de expertos *estúpidos* en dominios pequenos, pero non realizamos progresos reais cara á IA xeral

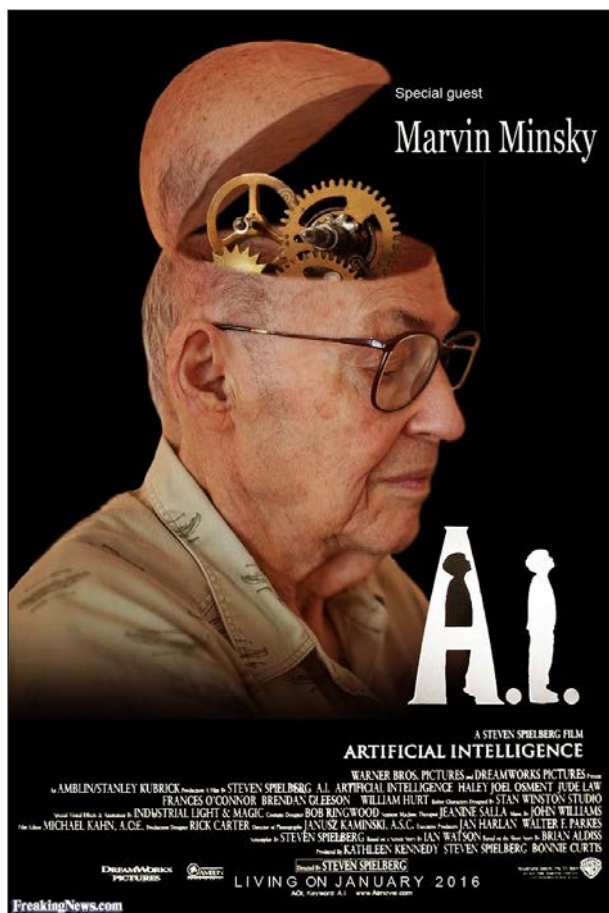
**Profesor Minsky, revisando a súa biografía, un sorpréndese unha chisca ao ver que vostede inventou un tipo novo de microscopio en 1957. En 1951 construíra a primeira rede neuronal artificial utilizando tres mil tubos de baleiro para simular a aprendizaxe Hebbià cunha rede de corenta neuronas. A invención do microscopio confocal tiña relación co seu interese en estudar como funciona o cerebro?**

De novo, cando era estudante, lin unha colección de artigos publicados no *Boletín de Biofísica Matemática* sobre modelos matemáticos das neuronas e as súas conexións. Un destes artigos era de Warren McCulloch e Walter Pitts, sobre como podían calcular e aprender as neuronas. McCulloch e Pitts baseaban o seu coñecemento da fisioloxía das neuronas na lóxica proposicional de Russell e na teoría da computa-

ción de Turing para modelizar as neuronas mediante unidades lóxicas. Gustoume ese artigo e comecei a interesarme pola cuestión de como funcionan realmente as neuronas e púxenme a traballar nas miñas propias teorías ao respecto. Un dos problemas era que cos microscopios dispoñibles daquela non había maneira de observar as neuronas cunha certa nitidez e, por ese motivo, a mediados da década dos cincuenta, botei anos traballando nun tipo novo de microscopio e inventei o confocal, que tiña mellor resolución e mellor contraste. Quitei algunhas fotos das neuronas, formulei algunhas teorías sobre como funcionan as neuronas e nunca máis volvíu utilizar o microscopio! Tantos anos de dedicación para unhas poucas fotografías!

**Pero continuou traballando nas súas teorías sobre como funcionan as neuronas, polo menos ata 1969, cando publicou o libro *Perceptrons*, con Seymour Papert.**

Si, nese momento pensei, igual que moitos outros, que o primeiro que cumpría facer era comprender como funcionan as neuronas e xa despois esclarecer progresivamente como funcionan partes cada vez maiores do cerebro, pero logo decateime de que as teorías de mediados dos cincuenta non eran mellores que as que xa había cen anos antes! Mesmo Freud, a finais do século XIX, tiña teorías moi-





Ramón López de Mántaras durante a entrevista con Marvin Minsky

to máis interesantes sobre como poden aprender as neuronas.

**E entón deixou de traballar nesta aproximación *bottom-up* da IA. Polo tanto, cal é a súa opinión sobre os grandes proxectos actuais que tratan de reproducir artificialmente, *bottom-up*, o cerebro, comezando polo nivel micro, como o proxecto *Human Brain* da UE?**

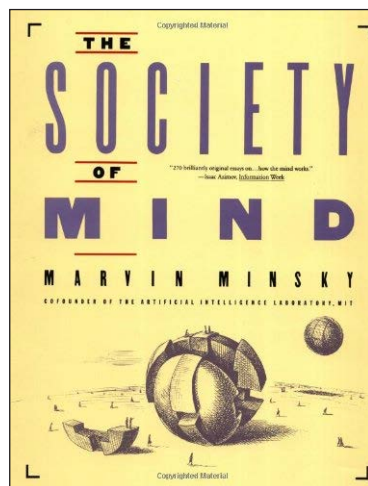
Son bastante escéptico. Eses proxectos non saben moi ben o que teñen que buscar no cerebro. O máis probable é que non atinen co que cómpre observar do cerebro co fin de entender como funciona. Temo que eses proxectos son unha perda de tempo e diñeiro. Sería mellor estudar e comprender primeiro o cerebro de animais moito máis simples antes que o dos seres humanos. Tamén sería mellor financiar mil proxectos cun millóns de dólares cada un, que dar mil millóns a un só proxecto.

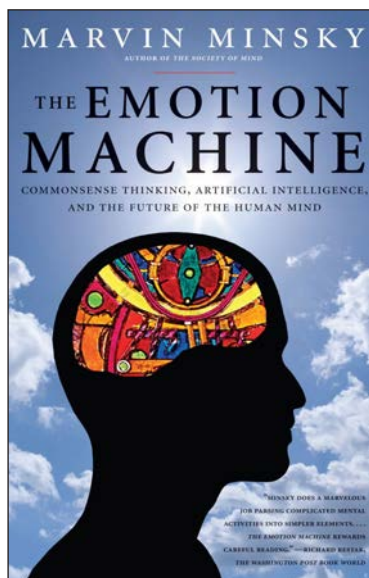
**De que traballo se sente máis orgulloso?**

Pois, se cadra, dun artigo que escribín en 1970 sobre diferentes liñas posibles de investigación en IA, titulado “Steps Towards Artificial Intelligence”. Creo que ese traballo predixo bastante ben o que moitos investigadores en IA fixeron durante os vinte anos seguintes. A finais dos anos setenta toda esa investigación comezou a decrecer, polo menos nos EUA, xa que o financiamento da investigación básica en IA diminuíu de xeito significativo e a xente dedicouse a desenvolver aplicacións.

**Refírese ao denominado *AI Winter*, que desprazou a comunidade de IA cara á construción de sistemas específicos e a esquecerse do obxectivo de construír intelixencias artificiais xerais?**

Si, por desgraza, a falta de financiamento en investigación básica contribuíu, entre outras causas,





ao denominado *AI Winter*. Durante máis ou menos as tres últimas décadas houbo pouco emprego interesante en IA debido a esta mudanza de obxectivos cara á construción de sistemas de IA demasiado específicos. Temos unha morea de expertos *estúpidos* en dominios pequenos, pero non realizamos progresos reais cara á IA xeral. Os avances máis interesantes en IA leváronse a cabo entre principios dos anos sesenta e finais dos setenta. As tendencias recentes en modelos probabilísticos de aprendizaxe non creo que sexan un avance significativo respecto ao que xa tiñamos hai cincuenta anos. Temos que abordar de novo as preguntas máis profundas da intelixencia xeral.

**Pero non lle parece que agora estamos asistindo a unha especie de rexurdimento do interese na IA xeral? E, en particular, na adquisición e representación do coñecemento de sentido común?**

Si, parece que hai máis persoas interesadas en problemas básicos como o do sentido común. Ese foi sempre un tema primordial en IA. Malia os éxitos recentes con Watson ou Siri, debemos ter claro que ningún destes sistemas saben que se pode lanzar un obxecto cunha corda

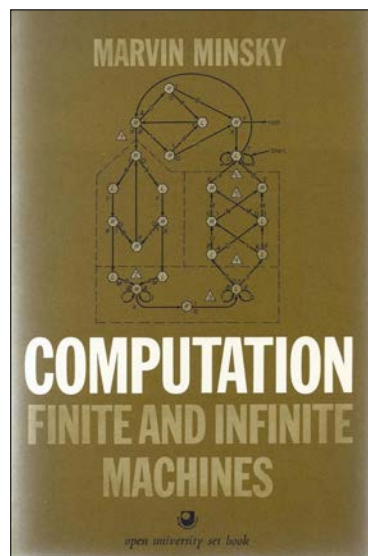
pero que non se pode empurrar. Por iso é necesario estudar o problema da aprendizaxe e a representación de coñecementos de sentido común.

**E o seu libro *The Society of Mind*? Non tivo máis impacto que "Steps Towards Artificial Intelligence"?**

Ese foi realmente o meu primeiro libro sobre IA e tiña capítulos cuxa lonxitude era dunha única páxina. Esa foi unha das razóns polas que tivo un grande impacto, pois a máis de curtos os capítulos eran bastante independentes entre si. Polo tanto, se o lector non entende ou non lle gusta un capítulo pódoo saltar sen que iso supoña un problema. Incluso moitos estudantes de secundaria son capaces de entender moitas teorías nese libro. O meu segundo libro, porén, *The Emotion Machine*, unha especie de segunda parte de *The Society of Mind*, tiña capítulos máis longos e teorías máis difíciles de entender e agora creo que iso foi un erro estratéxico. Talvez escriba outro libro simplificando as teorías e con capítulos máis curtos para que a xente as entenda.

**Sería fantástico! E se falamos agora un pouco sobre a música<sup>1</sup> e, en concreto, en relación co seu artigo de 1981 "Music, Mind, and Meaning"? Que o levou a escribir ese artigo?**

No ano 1981, organizaron no IRCAM unha conferencia titulada "The Composer and the Computer", e Tod Machover, que traballaba no IRCAM nese momento, convidoume a participar como un dos conferenciantes principais. Escribín a primeira versión de "Music, Mind, and Meaning" a partir desa conferencia. Nese artigo equiparo unha obra musical a un sistema de razoamento. Tomei como exemplo o primeiro movemento da *Quinta sinfonía* de Beethoven e expliquei como quería Beethoven que escoitáse-



mos esa obra, con cada frase fai-nos pensar nela mesma e na súa relación con todo o que a segue dunha maneira moi especial. De modo que a peza de seu é unha especie de *máquina* que nos ensina como temos que escoitar. Dito con outras palabras, argumento que escoitar música é unha experiencia de aprendizaxe. O artigo contén ideas sobre como funciona a música e a forma en que a escoitamos, o que significa para nós e como se relaciona coas emocións e a cognición humana en xeral e, de feito, moitas desas ideas relaciónanse tamén co que explico no libro *The Society of Mind*.

**Cuestións que aínda son problemas abertos, como o demostra o feito de que os estivemos discutindo onte mesmo durante o simposio!**

Si, moitas destas cuestións seguen sen resposta malia os avances recentes na neurociencia da música, así como na cognición e na percepción da música. O meu interese pola música, en relación co cerebro, débese ao feito de que a música ten moito que ver coas emocións e creo que a emoción é un tipo de razoamento. Ese é o motivo de que a música sexa importante para a IA. O meu libro *The Emotion*



*Machine* reflicte, claro está, a miña crenza de que a emoción é razoamento.

**Que opina da “Singularity”? Estará de acordo en que, tal como predixo Ray Kurzweil, pasará relativamente pronto?**

Hai avances, pero nos últimos dez anos, máis ou menos, retardáronse moito; polo tanto, son moito menos optimista que Ray Kurzweil sobre cando chegará.

**Ao meu entender, o máis probable é que non chegue nunca, pois a hipótese básica de que o coñecemento científico se está acelerando decontino non parece correcta, senón todo o contrario. A física, por exemplo, progresaba moito máis rapidamente hai cincuenta ou incluso cen anos que agora, non lle parece?**

Si, e iso é máis certo aínda no que se refire ao progreso da IA. Como xa dixen, retardouse en lugar de acelerarse. A razón é que os investigadores novos e brillantes non conseguen emprego para facer investigación básica. É difícil atopar oportunidades para facer investigación básica en IA. Así que, aínda que

me parece que a singularidade talvez sexa posible, eu non me preocuparía moito.

**Quen son os seus autores de ficción científica preferidos? Aprendeu algo da ficción científica?**

Aprendín moito da ficción científica. Un dos libros que máis me influíu foi un de relatos breves de H. G. Wells. Cada unha das historias tiña algunha hipótese boa, como por exemplo aquela en que alguén inventa a antigravidade, o que permitía a construción de naves espaciais de baixo custo e ir á Lúa e cousas polo estilo. Hipóteses que, mesmo non sendo factibles, abren portas ao pensamento e á imaxinación. Outro escritor que me gusta é Jules Verne. Tamén Aldous Huxley e moito máis adiante Isaac Asimov que, por certo, vivía preto da miña casa e, polo tanto, tivemos moitas ocasións de falar sobre robots, pero el sempre se negou a vir ao MIT para velos! Cando lle preguntei por que, díxome: “É que estou escribindo sobre robots do futuro intelixentes de verdade e ver os robots torpes que estades a construír podería

afectar a miña imaxinación”, e tiveren que admitir que levaba razón! Claro está, tamén me fixen amigo de Arthur Clarke e de Stanley Kubrick canda a miña relación como asesor de *2001 A Space Odyssey*. Arthur Clarke botou un par de semanas vivindo no piso de enriba desta casa onde estamos agora, e claro que falamos moito sobre ciencia e ficción científica. El era moi optimista en relación ao futuro da intelixencia artificial. Afirmaba que “calquera tecnoloxía avanzada dabondo é indistinguible da maxia” e máis de acordo eu non podería estar!

**É unha idea fantástica para rematar. Moitas grazas por atenderme! ■**

1. Margaret Minsky, filla de Marvin, escribiu: “Marvin ten a rara habilidade de improvisar fugas, pásao estupendamente e faino a miúdo. Foi un neno prodixio improvisando e compondo música; máis adiante, ao tempo que se formaba como científico, estudaba música en serio. Durante os anos sesenta e setenta gravou na casa moitas improvisacións contrapuntísticas e experimentaba con formas novas de tocar o piano, o instrumento que ama sobre todos os outros.

# Algunhas reflexións sobre o presente e futuro da Intelixencia Artificial

RAMÓN LÓPEZ DE MÁNTARAS BADIA

**O**bxectivo da Intelixencia Artificial (IA), lograr que unha máquina teña unha intelixencia de tipo xeral similar á humana, é un dos obxectivos máis ambiciosos que se propuxo a ciencia. Pola súa dificultade, é comparable a outros grandes obxectivos científicos, como explicar a orixe da vida, a orixe do universo ou coñecer a estrutura da materia.

Ao longo dos últimos séculos, este afán por construír máquinas intelixentes conduciunos a inventar modelos ou metáforas do cerebro humano. Por exemplo, no século XVII, Descartes preguntouse se un complexo sistema mecánico composto de engrenaxes, poleas e tubos podería, en principio, emular o pensamento. Dous séculos despois, a metáfora foron os sistemas telefónicos, pois parecía que as súas conexións se podían asemillar a unha rede neuronal.

Actualmente, o modelo dominante en IA é o modelo computacional baseado no computador dixital. Este artigo contén algunhas reflexións sobre a IA actual así como sobre o seu futuro.

## A hipótese do Sistema de Símbolos Físicos e a IA: IA débil *versus* IA forte

Centrándonos na IA actual, é importante distinguir entre IA débil e IA forte, dúas visións que se corresponden, respectivamente, cos dous seguintes intentos de definición:

Definición 1 (IA débil) A IA é a ciencia e a enxeñería que permite deseñar e programar ordenadores de forma que realicen tarefas que requiren intelixencia.

Definición 2 (IA forte) A IA é a ciencia e a enxeñería que permitirá replicar a intelixencia humana mediante máquinas.

Con motivo da recepción do prestixioso Premio Turing en 1975, Allen Newell e Herbert Simon formularon a hipótese do Sistema de Símbolos Físicos (SSF) segundo a cal “todo sistema de símbolos físicos posúe os medios necesarios e suficientes para levar a cabo accións intelixentes”. E dado que os seres humanos somos capaces de mostrar condutas intelixentes no sentido xeral, entón, de acordo coa hipótese, nós somos tamén sistemas de símbolos físicos.

Convén aclarar a que se refiren Newell e Simon cando falan de “sistema de símbolos físicos”. Un SSF consiste nun conxunto de entidades denominadas símbolos que, mediante relacións, poden ser combinadas formando estruturas máis grandes (como os átomos que se combinan formando moléculas) e que poden ser transformadas aplicando un conxunto de procesos.

Estes procesos poden crear novos símbolos, crear e modificar relacións entre símbolos, almacenar símbolos, comparar se dous símbolos son iguais ou distintos, etc. Estes símbolos son “físicos” en tanto que teñen un substrato físico-electrónico (no caso dos computadores) ou físico-químico-biolóxico (no caso dos seres humanos).

Efectivamente, no caso dos computadores os símbolos realízanse mediante circuitos electrónicos dixitais e no caso dos seres humanos mediante redes de neuronas.

En definitiva, consonte coa hipótese SSF, a natureza do substrato (circuitos electrónicos ou redes neuronais) carece de importancia a condición de que o devandito substrato permita procesar símbolos. Non esquezamos que se trata dunha hipótese e, por tanto, non debe de ser nin aceptada nin rexeitada *a priori*. En calquera caso,