Intelixencia artificial: o noso último invento[1]

Senén Barro Ameneiro

Apertura do curso académico da RAGC 2023, Santiago de Compostela, 16 de xaneiro de 2023

Presidente da RAGC, autoridades, académicas e académicos, familia, amigas e amigos, moitas grazas pola vosa presenza.

É un pracer e unha honra que me corresponda impartir a lección coa que se inaugura un novo curso académico da RAGC.

Parabéns aos membros da nova comisión executiva da RAGC, e agradecemento aos anteriores membros de dita comisión polo excelente traballo desenvolvido.

Resumo

A intelixencia artificial ou IA, como adoita chamárselle tamén, no só é un dos campos científico-tecnolóxicos máis activos e cun maior impacto social e económico, senón que está de moda. A *Fundación del Español Urgente*, FundéuRAE, acaba de elixir "intelixencia artificial" como palabra do ano 2022. Realmente como expresión do ano, xa que son dúas palabras e non unha.

Aínda que a intelixencia artificial comezou con intentos de modelar o pensamento humano como vía para avanzar tanto no coñecemento do noso cerebro como para desenvolver máquinas capaces de imitalo, a día de hoxe, despois de sucesivos períodos de luces e sombras, vivimos unha aproximación moito máis pragmática, que busca desenvolver sistemas que empregan a IA na resolución de problemas. Se ben non temos aínda unha forma de cuantificar ou medir a intelixencia dun dispositivo, en xeral podemos dicir que unha máquina ou programa informático posúe intelixencia [artificial] cando dito sistema ten unha autonomía e riqueza de comportamento significativas en dominios dinámicos e complexos -un coche autónomo, por exemplo-, é capaz de aprender a través da experiencia adquirida sobre a contorna na que opera -un programa que mellora a medida que xoga ao xadrez ou que identifica os nosos gustos literarios mentres navegamos pola rede- e/ou presenta un alto grao de competencia en áreas especializadas do coñecemento humano -un sistema experto nun dominio médico ou para a concesión de créditos bancarios, poñamos por caso. A través da IA as máquinas xa son capaces de superarnos en moitos dominios ou tarefas concretas, non só físicas, senón tamén, e cada vez máis, as que requiren competencias cognitivas, como no recoñecemento de imaxes, no diagnóstico baseado en imaxes médicas ou sinais fisiolóxicas, na tradución entre linguas… e xa non digamos no xadrez ou calquera outro xogo de estratexia.

---

[1] Aínda que este texto foi especificamente elaborado para a ocasión, contén abundantes referencias e incluso partes de outros escritos previos, nomeadamente artigos elaborados para medios de comunicación. En todo caso, tratase sempre de escritos propios.

A intelixencia artificial, está na fala da xente e déixanos pampos, un día si e outro tamén, cos logros aos que dá lugar en case calquera campo da actividade humana. Interesa moito socialmente porque as súas capacidades son tantas, moitas aínda por construír e case todas por imaxinar, que levanta enormes expectativas e non poucas preocupacións. Pode que por iso sexan tantos os paradoxos, mitos, malentendidos, medias verdades e incluso mentiras como puños, que escoitamos e lemos en relación á IA. Precisamente esta lección vai virar ao redor dalgúns deles, para tentar explicar dun modo tamén entretido, e non só formativo, o que é a IA, de onde vén, onde está e, finalmente, especular co feito de que poida chegar a ser o noso último invento. Que chegue a selo, e sexa para ben, dependerá de como o fagamos polo camiño nós, as persoas.

A IA está na terceira idade

Un equívoco frecuente é pensar que o ámbito da intelixencia artificial é recente. Nin moito menos, salvo que case sete décadas de vida intensa parézannos un case nada.

Nun documento presentado á Fundación Rockefeller, datado o 31 de agosto de 1955, un grupo de científicos, encabezados por John McCarthy, solicitaba fondos para organizar un encontro dun par de meses, durante o verán de 1956, en Dartmouth College, unha universidade privada situada en Hanover, Novo Hampshire, Estados Unidos. No documento xa foi utilizado o termo "intelixencia artificial", e conxecturábase sobre a posibilidade de que cada aspecto da aprendizaxe ou calquera outra característica da intelixencia puidese, en principio, ser descrita con tanta precisión que permitise ser simulada por unha máquina. O nome do campo, por certo, resultou todo un acerto, por moito que poda discutirse se é científica e lingüisticamente preciso[2].

Mesmo antes desa data, e de acuñar o nome de intelixencia artificial para este incipiente campo, producíronse algunhas das contribucións máis relevantes desta nova disciplina do saber científico e do desenvolvemento tecnolóxico. Non procede facer aquí unha revisión deses antecedentes, pero polo menos permítanme citar un par de exemplos. Por unha banda a proposta do primeiro modelo matemático de neurona artificial foi realizada por Warren McCulloch e Walter Pitts en 1943. Tratábase dun modelo moi elemental de neurona artificial, capaz en todo caso de sintetizar, convenientemente combinadas en redes, todos os conectores lóxicos e calquera función calculable. McCulloch e Pitts uniron baixo a súa proposta de neurona formal a Russell e a lóxica de proposicións, con Turing e a súa máquina de computación, e Sherrington, e a súa teoría das sinapses neuronais. O seu artigo foi moi influente, ata o punto de que von Neumann cambiou o sistema decimal polo binario na súa estratexia de deseño de computadoras. A pesar da súa extrema sinxeleza, este modelo de neurona segue sendo o referente dos modelos actuais de computación neuronal.

[2] El concepto de Inteligencia Artificial es una prosopopeya -personificación o atribución de cualidades humanas a seres inanimados o animales-, ya que en parte se identifica con la capacidad de las máquinas para hacer aquello que nosotros hacemos y que, de algún modo, atribuimos a nuestra inteligencia, la humana. Pero quién sabe, quizás algún día el significado de prosopopeya haya de reescribirse como la atribución de cualidades humanas o de máquinas inteligentes a seres, animados o no, que carezcan de dicha inteligencia.

Podemos dicir, por tanto, que a intelixencia artificial naceu na práctica bastante antes de nacer formalmente, e fíxoo bioinspirándose no elemento anatómico e funcional básico do noso cerebro e o doutros seres vivos: a neurona.

Outra contribución seminal tivo lugar en 1950, cun artigo de Alan Turing, unha das mentes máis brillantes do século XX. Turing propuxo un "xogo da imitación", que logo se popularizaría como test de Turing, nunha publicación na revista *Mind*, titulada: "Computing Machinery and Intelligence". O artigo comeza así: "Propoño considerar a seguinte cuestión: Poden pensar as máquinas?". Basicamente consiste no seguinte: unha persoa, á que denominaremos xuíz, dialoga con outra persoa e unha computadora, aos que non pode ver, tentando pescudar quen é quen. Ademais, o diálogo realízase a través dun medio que non delate sen máis á persoa ou á computadora. Imaxinemos, por exemplo, que utilizamos para iso un teclado e unha pantalla. Se a través da conversación o xuíz é incapaz de discriminar a persoa da máquina nun período de tempo razoable, dise que a computadora superou o test e pode ser considerada intelixente. Polo menos así o pensaba Turing.

Máis de sete décadas despois, chatbots como ChatGPT fan pensar que o test de Turing superouse, ou polo menos que non estamos moi lonxe de logralo. Todo depende do estritos que nos poñamos coas regras neste xogo no que poñemos a máquinas a facerse pasar por persoas. En todo caso, o que se está evidenciando é que as capacidades do cerebro humano están moi por encima das dunha máquina que saiba moito e saiba dialogar en linguaxe natural.

As neuronas artificiais non son como as nosas

Nos anos 50 e 60, a investigación en IA centrouse principalmente nas Redes Neuronais Artificiais (RNA). Posteriormente chegaron as solucións baseadas na representación explícita do coñecemento potencialmente útil para resolver problemas e a aplicación de mecanismos de razoamento automático para aplicalo aos problemas antes mencionados. Este enfoque, propio dos denominados Sistemas Baseados en Coñecemento (SBC), prevaleceu en boa medida ata finais do século pasado.

Aparentemente debería ser fácil deseñar sistemas baseados no coñecemento que as persoas fomos acumulando ao longo de séculos de continuos avances no saber e na súa aplicación. Pero nada máis lonxe da verdade. Por exemplo, a dificultade para unha análise introspectivo das nosas decisións e accións é tamén unha dificultade para poder avanzar no deseño de máquinas que poidan imitalas.[3]

[3] Michael Polanyi foi un brillante filósofo, pero tamén cultivou os campos da economía, a medicina e a físico-química. Nun dos seus libros, "A dimensión tácita", publicado en 1966, deixaba claro que as persoas sabemos máis do que podemos expresar. Exemplos evidentes e cotiáns son a pericia dun taxista, que non pode ser adquirida en clases de teoría na autoescola nin ser explicada polo propio taxista. Tampouco podemos explicar como recoñecemos a cara dun amigo de lonxe e ata detrás dunha máscara. Podemos explicar como somos capaces de entender a nosa lingua materna ou aprender un novo idioma?
O paradoxo de Polanyi convértese así nunha limitación para o deseño de máquinas que poidan alcanzar ou mesmo superar a nosa pericia na resolución de problemas complexos en dominios especializados do coñecemento. Unha forma de superala, ou polo menos de circunvalala, é a aprendizaxe automática. Se as

Pensemos, por exemplo, no deseño de sistemas expertos baseados no coñecemento de quen é especialistas nun ámbito do saber. Persoas que cada día fan o seu traballo cunha profesionalidade e un coñecemento extraordinarios, como os especialistas en calquera campo da medicina, non son capaces de explicar unha boa parte de como o fan e cústanos case a vida "extraerlles" o coñecemento necesario para tentar representalo computacionalmente e utilizalo en procesos de razoamento artificial. Seino por experiencia. De feito, iniciei a miña carreira como investigador no ámbito da IA aplicada á medicina. Vivín de primeira man as dificultades de deseñar sistemas expertos, baseados no coñecemento dos médicos, neste caso. Esa foi unha das razóns de que explorase en paralelo a computación neuronal e a aprendizaxe automática, naquel momento, finais dos 80 e principios dos 90, escasamente desenvolvidos.

Aínda que a finais dos oitenta producíronse algúns resultados de investigación significativos en computación neuronal que devolveron o interese polas Redes Neuronais Artificiais, foi a principios da pasada década cando estas reapareceron con máis forza que nunca. Durante anos houbo unha intensa procura para evitar os principais problemas dos SBC, sendo a aprendizaxe automática o enfoque máis interesante e potencialmente útil. Desde o cambio de século a aprendizaxe automática ha ida gañando terreo no campo da IA, debido ao deseño de novos algoritmos, como os baseados na aprendizaxe profunda, o aumento constante da potencia de cálculo e a existencia de inmensos bancos de datos dos que aprender. Todo iso permitiu avances espectaculares para acometer problemas moi complexos, aínda que tamén moi específicos. Por exemplo, en 2012 o equipo do profesor Geoffrey Hinton, da Universidade de Toronto, que gañou conxuntamente con Yoshua Bengio e Yann LeCun o premio Turing[4] en 2018, esmagou a todos os seus rivais no ImageNet Large-Scale Visual Recognition Challenge. Aplicando o seu algoritmo "SuperVision", baseado en "redes neuronais convolucionais profundas", obtiveron un erro do 16,4% na clasificación de 1,2 millóns de imaxes de alta resolución pertencentes a mil clases diferentes. SuperVision abriu o camiño para superar aos humanos en tarefas de recoñecemento de imaxes e outros problemas igualmente complexos, como o recoñecemento de voz, a xeración automática de texto, os sistemas de diálogo, e un etcétera que cada día faise máis grande. En calquera caso, o gran reto da IA segue sendo conseguir unha IA de propósito xeral, como a nosa intelixencia; un reto que estou seguro de que tamén requirirá de mecanismos xerais de aprendizaxe, algo llo que estamos moi lonxe.

En todo caso, e por moito que nos estea sorprendendo a intelixencia artificial baseada en hipercomplexos modelos de computación neuronal, non podemos comparar as súas

máquinas poden aprender a partir de exemplos, o que adoitamos chamar datos, poderemos non depender, ou non tanto, do coñecemento humano ou do modo humano de resolver os problemas. Por exemplo, poderemos deseñar un sistema para a detección de nódulos cancerosos a partir de radiografías de tórax. Necesitamos, iso si, un bo número de radiografías sobre as que se identificaron previamente, e con suficiente precisión, os nódulos malignos. Se os exemplos son bos, numerosos e suficientemente representativos do problema que pretende aprenderse, as cousas adoitan funcionar ben. No caso comentado, superando incluso a capacidade de recoñecemento visual do especialista en radioloxía.

[4] O Premio Turing está considerado como o Premio Nobel das ciencias da computación.

capacidades coas dun cerebro humano. De feito, os modelos de neurona artificial que se usan na computación neuronal son a unha neurona real o que unha lente a un ollo.

Incluso as máis sofisticadas redes neuronais artificiais non son máis que un remedo das neuronas reais, polo menos no coñecido. Poñamos por caso os procesos de retropropagación utilizados en moitas arquitecturas de computación neuronal. Dita retropropagación non parece ter correlación no noso cerebro, xa que as sinapses químicas son unidireccionais –non asi as eléctricas-. Operan cara a adiante, por dicilo así. Si existen no noso cerebro circuítos neuronais con realimentación, pero as conexións teñen un sentido único. Non hai por tanto retropropagación de sinais que poidan asemellarse aos algoritmos de retropropagación empregados frecuentemente en RNA. Como non hai tampouco nos modelos de neurona artificial nada semellante ás espiñas dendríticas ou aos neuromoduladores. Ademais, a anatomía non o é todo, senón que se comprobou que a plasticidade neuronal vai máis aló das conexións, podéndose alterar o funcionamento dos circuítos neuronais mediante compostos químicos que poden "navegar" polo noso cerebro.

Ás máquinas dáselles mellor facer a selectividade que unha tortilla de patacas

A pesar dos incribles avances da IA. aínda a maior parte do que facemos as persoas, en particular algunhas das cousas que consideramos máis sinxelas, non poden facelas as máquinas. Si, é certo que poden aspirar a casa sen cansarse, pero só se o espazo é plano e non hai moitas complicacións por medio. Desde logo, non esperemos que de momento levanten as alfombras. Poden diagnosticar nódulos cancerosos en radiografías de tórax mellor que calquera especialista en radioloxía, pero non poden manter unha simple conversación co paciente sobre o seu estado de ánimo. Poden guiarnos paso a paso para cociñar miles de receitas, informándonos de todo o coñecido sobre os ingredientes, pero non se manexan nunha cociña normal para facer unha tortilla de patacas.

O que para as persoas é máis difícil e consómenos moito tempo e enerxía mental, como os procesos de razoamento lóxico, facer cálculos matemáticos ou tratar con enormes cantidades de datos ata extraer deles o que nós nin sequera sospeitamos que existe, é o máis fácil de levar ao ámbito da computación. O tándem algoritmo-computadora permite ir máis aló do que está ao noso alcance nestes casos. Pola contra, hai moitas cousas triviais para nós que as máquinas resolven mal ou nin sequera fan a día de hoxe, como recoñecer unha escena nun dicir amén, e nunca mellor dito, ou manter unha conversación de cafetería. A isto púxoselle nome: paradoxo de Moravec, á que deu nome un investigador no campo da robótica, que xa nos anos 80 destacou esta aparente contradición.

A contradición só é aparente, en todo caso. O noso cerebro non evolucionou para ser unha boa calculadora nin un bo razoador lóxico, pero na medida en que puidemos detallar suficientemente ben como se poden sistematizar estas tarefas, resultounos razoablemente fácil automatizalas mediante computadoras. Facendo unha analoxía co mundo da computación, son esas tarefas para as que fomos desenvolvendo "programas mentais", que executamos na nosa máquina de propósito xeral, o cerebro, e que podemos trasladar a outras máquinas, neste caso artificiais, máis eficientes para este tipo de cousas. Pola

contra, a forma na que percibimos o mundo, falamos, camiñamos, manipulamos obxectos… é algo que se foi asentado no noso cerebro ao longo de toda a evolución, non algo que realizamos mediante procesos conscientes e intencionados pola nosa banda. Aprendémolo, digámolo así, de modo natural, sen darnos conta, e apenas sabemos como o facemos.

Isto explica que para unha máquina sexa máis fácil aprobar a selectividade –xa hai case unha década que un software xaponés logrou superar os exames de ingreso cunha cualificación que lle permitise o acceso a case todas as universidades privadas do país- ou mesmo o MIR, que poder manipular a contorna como o fai un neno de dous anos. Quizais isto debería facernos pensar se certo tipo de exames teñen sentido, cando o que queremos realmente é que as persoas fagan sobre todo aquilo que aínda non fan as máquinas.

A evolución das máquinas é moito máis rápida que a nosa

En 1958 o New York Times publicou un artigo titulado: "Un novo dispositivo da Mariña aprende coa práctica". Este xornal facíase eco dunha rolda de prensa do prestixioso psicólogo estadounidense, Frank Rosenblatt. O artigo comezaba sinalando que se estaba desenvolvendo un computador electrónico -naquel momento isto era en si mesmo un gran avance da tecnoloxía de computadoras-, que sería capaz de camiñar, falar, ver, escribir, reproducirse e ser consciente da súa existencia.

Como adoita ocorrer cando falamos da tecnoloxía, en particular da vangarda tecnolóxica, houbo entón un exceso de optimismo sobre o que en poucos anos esperábase conseguir. A coñecida como Lei de Amase dinos que tendemos a sobreestimar os efectos dunha tecnoloxía no curto prazo e a subestimalos no longo prazo.

En todo caso, aínda que os avances da IA non foron tan rápidos como se pensaba naquel momento, co tempo as máquinas foron logrando case todo o que se aventuraba no mencionado artigo do New York Times. En 2016, por exemplo, as máquinas xa nos superaron no recoñecemento de imaxes e hoxe son máis competentes que os dermatólogos en discriminar os distintos tipos de cancro de pel. Tamén nos superan no recoñecemento de voz e en comprensión de texto, aínda que non entendan o significado real do que recoñecen. Tampouco comprenden o significado do que traducen, o que non impediu que en 2018 alcanzasen o nivel de competencia humana na tradución entre inglés e chinés. E si, as máquinas saben escribir e xa son moitos os medios de comunicación que utilizan programas de redacción automática de noticias, indistinguibles en moitos casos das que podería redactar unha persoa. É máis, o caso de ChatGPT, o chatbot que nos asombrou a todos a finais de 2022, é quizais o mellor exemplo dos avances espectaculares e acelerados aos que estamos a asistir.

Unha máquina deséñase cun propósito e cada unha dos seus partes responde a un obxectivo e realiza unha función no conxunto dun sistema máis ou menos complexo. Pode non ser un deseño óptimo para o que se busca, pero responde a un propósito claro desde o seu mesma concepción. Por iso os avances do sintético son ás veces tan rápidos e espectaculares. Ademais, no mundo da computación e do computable cada avance serve para camiñar cara a logros aínda maiores. As computadoras de onte permitiron chegar ás

computadoras de hoxe e estas son os que farán posible as do día de mañá, e tamén todo o que se construirá con elas. A evolución, pola contra, necesita miles e miles de anos para que un ser vivo sexa apreciablemente máis intelixente que os seus predecesores. Por iso cabe pensar que a intelixencia de persoas e máquinas chegará a ser comparable algún día. Iso si, salvo que antes nos evaporemos, como a auga que finalmente non chega ao mar.

A nosa intelixencia é de propósito xeral e a artificial non

Unha intelixencia artificial a nivel humano, ou IA forte, sería aquela que tivese capacidades comparables á nosa en todos os sentidos, incluída a consciencia ou a intelixencia emocional e social. Esta preséntase normalmente como contraposición á IA de propósito específico, ou débil, que é a actual, de feito, preocupada por solucionar problemas concretos e non por conseguir unha intelixencia artificial humanizada.

Aínda que ás veces identifícanse os termos de intelixencia artificial xeneral e IA forte, esta debe mellor reservarse para a IA a nivel humano. A IA xeral, ou de propósito xeral, céntrase sobre todo en aspectos conductuais, en capacidades xerais e moi versátiles para a aprendizaxe e a resolución de problemas, como os nosos, pero non ten que asociarse necesariamente con todo o que identificamos co noso cerebro e os seus procesos cognitivos, ou coa propia autoconsciencia. Se é posible aquilo sen isto, iso si, xa é fariña doutro costal.

En todo caso, estamos moi, pero que moi lonxe dunha intelixencia artificial que se nos pareza. Non temos nin idea de como se produce a nosa consciencia; non sabemos de certo se gozamos do libre albedrío ou se é só un bonito desexo; aínda que podemos simular algo que remede algunhas emocións, non é máis que case nada de momento. Pero non é o único que segue fóra do alcance das máquinas. Tamén o humor é algo xenuinamente humano e a creatividade humana dista moito da incipiente, aínda que crecente, creatividade maquinal.

A historia da IA está chea de predicións sobre cando se alcanzará unha intelixencia artificial de propósito xeral, ou mesmo unha intelixencia comparable ou ata superior á nosa. Por exemplo, Herbert Simon predixo en 1965 que as máquinas serían capaces en vinte anos de facer calquera traballo que puidese facer unha persoa; Marvin Minsky afirmou en 1967 que "dentro dunha xeración... o problema de crear intelixencia artificial estará substancialmente resolvido"; e Raymond Kurzweil, director de enxeñería en Google, predixo en 2005 que a IA forte farase realidade en 2045. Quédannos algo máis de dúas décadas ata entón, pero parece pouco tempo para iso, sobre todo porque non temos nin a menor idea de que camiño percorrer.

O obter unha intelixencia artificial de propósito xeral non só é un reto científico-tecnolóxico, senón unha necesidade para ter máquinas que realmente poidan resolver todo tipo de problemas e non só algúns e un a un, como ocorre hoxe día. Neste momento cada problema que é abordado a través da IA require dun deseño específico, lento, custoso e pouco menos que inútil para acometer a partir del outros problemas distintos. A mellor solución para o xogo do xadrez non servirá en absoluto para xogar ao tres en raia. É verdade que hai matices a esta afirmación, e que é posible deseñar modelos neuronais en particular con capacidades interesantes para máis dun problema -para unha tipoloxía de

problemas, mesmo-, que despois poden ser adaptados cun menor esforzo á resolución de cada problema específico -no meu grupo de investigación esta é unha das liñas abertas, en particular no ámbito da robótica-, pero non é isto no que pensamos cando falamos dunha intelixencia de propósito xeral, nin moito menos.

A intelixencia e a infalibilidade non se levan ben

A intelixencia, sexa natural ou non, non garante cero erros. De feito, a infalibilidade e a intelixencia son incompatibles. Podo garantir que unha computadora cometerá cero erros multiplicando dous matrices, pero non guiando un coche.

En todo caso, sexa pola incompatibilidade entre infalibilidade e incremento na intelixencia, ou pola autonomía que poida alcanzar unha intelixencia artificial moi desenvolvida, é lóxica nosa preocupación sobre o que poidan facer as máquinas intelixentes pola súa conta, pero ao noso risco. Cada vez deixamos máis responsabilidades nas súas "mans". Naqueles casos nos que a máquina aprende a mellorar a súa desempeño e faio ademais en condicións reais de uso,[5] non é posible testar de antemán todos os detalles do seu funcionamento. Isto pode levar a comportamentos anómalos, aínda que os obxectivos fixados e consígnalas dadas para a súa consecución sexan claros e virtuosos. Unha máquina programada para defender custe o que custe a súa "supervivencia" podería actuar contra aquelas persoas que tentasen desconectala, como ocorre con HAL, a computadora de "2001: Unha odisea do espazo", a película de Stanley Kubrik que xa cumpriu o seu medio século de vida, e aínda segue con moi boa saúde.

Do mesmo xeito, a infalibilidade dunha intelixencia artificial á hora de conseguir un obxectivo podería ter tamén consecuencias desastrosas para nós. Pensemos nunha intelixencia artificial deseñada para acabar coa pandemia se decidise aniquilar á especie humana. Cumpriría sen dúbida o encargo, pero dun modo francamente indesexable. É un caso extremo e imposible hoxe día, pero que ocorrería se unha arma intelixente é deseñada para inflixir o maior dano posible ao inimigo co menor número posible de vítimas humanas? En principio parece unha boa consigna, pero no empeño por cumprir escrupulosamente coas ordes recibidas, a arma podería decidir acabar coa vida do responsable do seu deseño se descubrise que este estaba a ser espiado polo inimigo. As armas intelixentes tamén poden disparar pola culata.

[5] Unha das liñas de investigación que estamos a desenvolver actualmente no meu grupo é a aprendizaxe federada, unha técnica de aprendizaxe automática no que cooperan múltiples dispositivos ou computadoras, chamémoslle clientes, nos que cada un opera sobre os seus datos locais, sen intercambialos. A cooperación permite construír un modelo de aprendizaxe automática común, que en todo caso pode ser adaptado ás circunstancias singulares de cada cliente. A aprendizaxe federada foi proposta en 2016, fundamentalmente para abordar cuestións críticas nalgunhas aplicacións reais, como a privacidade ou a seguridade dos datos, pero as súas posibilidades son moitas máis, e están en boa medida por explorar e desenvolver. Por exemplo, a rapidez e robustez no proceso de aprendizaxe automática. No grupo estamos a investigar, entre outras cousas, na aprendizaxe federada continua, necesario en contornas dinámicas, onde a contorna local no que opera cada cliente é cambiante.

As máquinas pensan e comprenden, pero non como nós

Segundo o dicionario da Real Academia Española, pensar ten entre as súas acepcións as de: formar ou combinar ideas ou xuízos na mente; examinar mentalmente algo con atención para formar xuízo; tamén formar na mente un xuízo ou opinión sobre algo. Todo iso pode facelo en certo xeito unha intelixencia artificial, se omitimos, está claro, a referencia á mente, que o mesmo dicionario define, entre outras posibles acepcións, como: "conxunto de actividades e procesos psíquicos conscientes e inconscientes, especialmente de carácter cognitivo".

Por tanto, poderiamos afirmar que un sistema experto pensa cando diagnostica unha enfermidade cardiovascular ou ao aconsellar a concesión dun crédito bancario. Tamén o fai un programa informático baseado en lóxica cando demostra un teorema e un robot autónomo cando decide como evitar un obstáculo que aparece no seu camiño. O que está claro é que a forma de lograr que as máquinas pensen e sexan intelixentes non ten que seguir cos ollos pechados o modo en que o facemos as persoas. Entre outras cousas porque isto ignorámolo en boa medida.

Por outra banda, que é comprender? Un mero sintetizador de texto a voz non comprende o que le, como tampouco nós faciámolo de nenos moitas veces cando o profesor mandábanos ler no alto e os nervios e as présas impedíannos ir pensando no que liamos. Poñamos outro exemplo. Se tivésemos a unha persoa sen ningún tipo de interacción co mundo, máis aló de ensinarlle unha linguaxe básica co que puidese construír frases gramaticalmente correctas, esa persoa non entendería o significado daquilo que di ou escoita. O significado dáo o poder asociar aquilo que pensamos, que dicimos, que escoitamos, que lembramos... cunha realidade mundana ou abstracta, pero realidade en todo caso. Cando dicimos que entendemos a palabra "can", non é por saber recoñecer nela as letras que a forman e a pertenza da mesma a un dicionario no que está descrito o seu significado, senón polo feito de asociala inmediatamente a algo que coñecemos ou, en todo caso, por poder facelo a través da súa descrición. Se alguén nos di que outra persoa adoptou unha actitude circunspecta e non sabemos o que iso significa poderemos resolver a nosa ignorancia sobre o termo circunspecto buscando unha definición que si teña para nós correlación co que coñecemos. Dun modo semellante pode comportarse unha máquina e, na medida en que os símbolos que manexe poidan ser correlacionados con aquilo que "coñece" e sírvalle para manexarse con competencia no mundo, polo menos no mundo no que opera, esa máquina estaría a entender ditos símbolos. Enténdeos despois de que sexa capaz de usalos con competencia nun mundo real ou imaxinario, segundo o caso.

Dito isto, se un sistema é capaz de razoar sobre un texto, resúmeo, pode contestar a preguntas sobre o que nel dise… si comprenderá en certo xeito. Probaron ChatGPT? Dirían que en xeral comprenden o que vostedes lle escriben?

En definitiva, se lemos un texto simplemente facendo a conversión do texto a voz, pero non asociamos o lido co xa coñecido, non incorporamos ao noso saber nada do lido, non o usamos con ningún outro fin que o da lectura, nin no momento nin no futuro, poderemos dicir que non comprendemos o lido e que si o fixemos en caso contrario. O mesmo ocorrerá coas máquinas. Se unha máquina pode ler textos da internet e establecer

relacións entre o que le e o que sabe do mundo, de modo que poida usalas dalgún modo nese momento ou no futuro, estará en maior ou menor grao comprendendo o lido. Por exemplo, se le que os polbos poden usar o seu rádula para perforar a cuncha dunha ameixa á altura á que se atopa no seu interior o músculo adutor deste molusco, o que lle permite pechar con forza as súas valvas, e despois ve a fotografía dunha cuncha delicadamente perforada, podería asociar ambas as cousas para concluír que é probable que o buraco sobre a cuncha fose causado por un polbo, que, despois, seguramente lla comeu.

Ese "comprender" ten un enorme valor no contexto da intelixencia humana e tamén o terá no da IA. Por exemplo, facilita a explicación (supoño que é un lobo, xa que está nun bosque nevado, podo dicir ao ver unha fotografía), o razoamento e a aprendizaxe (aínda que leva un colar de can e vai atado a unha correa, parece un porco e como se que hai un aumento no número de porcos criados como mascotas, efectivamente podería ser un porco a pesar de levar colar; ademais, se así o asumo, poderei identificar máis facilmente outros porcos semellantes, aínda que non leven colar nin vaian atados por unha correa); e a creatividade mesma (ten un buraco, polo que podería meterse un fío e vin colares feitos de sementes con buracos polos que pasa un fío, logo… podería facer un colar con estas cunchas perforadas).

A intelixencia artificial e a creatividade

En outubro de 2018 a sala Christie´s en Nova York vendeu o cadro "Edmond de Belamy" por 432.500 dólares. Non é un prezo moi elevado para o que alí adoita poxarse, pero non está nada mal para unha obra creada por unha intelixencia artificial. De feito, foi a primeira obra realizada por unha IA que foi vendida nunha casa de poxas. Os seus "autores humanos" son o grupo francés que se fai chamar "Obvious", tres novos de París sen formación artística, pero que se consideran a si mesmos artistas e que buscan, como eles mesmos din, democratizar a arte coa axuda da IA. Para facer esa obra adestraron unha rede neuronal cunha selección de cadros clásicos, de modo que a rede aprendeu a xerar dixitalmente obras novas baseándose nos exemplos aprendidos. A partir de aí, como eles mesmos contaron: "seleccionamos as imaxes que máis nos gustaron, imprimímolas, enmarcámolas nun cadro dourado e puxemos prezo". E non lles foi mal, certamente.

Hai pouco un home chamado Jason Allen gañou o primeiro premio na categoría de arte dixital do concurso de belas artes da Feira Estatal de Colorado, en EE. UU. A obra, titulada: "Théâtre D'opéra Spatial", é realmente fermosa, de espléndida factura e moi orixinal na súa composición. É mesmo engaiolante, ou polo menos a min pareceumo, a pesar de que a vin só na pantalla do meu ordenador e non no lenzo no que Allen imprimiuna.

Puxen un par de exemplos de cadros creados por intelixencia artificial, pero a creatividade das máquinas non se restrinxe ao ámbito do que denominamos arte. A creatividade, ese producir algo novo, está moi presente xa na xeración automática de código, no deseño de novos dispositivos, como antenas de telefonía, de campañas de márketing, de novos deseños para produtos xa existentes ou mesmo de produtos completamente novos.

De feito, podemos dicir que o ano que acabamos de deixar atrás, 2022, foi o da IA creativa. A denominada IA xenerativa volveu creativa á IA, e isto non fixo máis que

empezar. Pareceunos sorprendente a capacidade de xerar rostros hiperrealistas ou reproducir a voz doutra persoa con enorme realismo tamén, e así poder poñerlle na súa boca o que nunca dixo. Pero isto nada ten que ver coas posibilidades que permiten intelixencias artificiais como DALL-E 2, baseado en aprendizaxe profunda, como case todas estas ferramentas, e que é capaz de producir imaxes a partir de descricións de texto; ou as ferramentas de Google e Meta para facer o propio, pero xerando vídeos. Por suposto, tamén ChatGPT, neste caso como chatbot capaz de dialogar con persoas cunha soltura e un coñecemento absolutamente inéditos e sorprendentes para unha intelixencia artificial.

Ningunha intelixencia é imprescindible

John Maynard Keynes acuñou a expresión "desemprego tecnolóxico" en 1930. Nun ensaio especulativo titulado "Posibilidades económicas para os nosos netos", predixo que o mundo estaba ao bordo dunha revolución no que se refire á velocidade, a eficacia e o "esforzo humano" nunha gran variedade de industrias. "Estamos a sufrir unha nova enfermidade da que algúns lectores quizá non oísen aínda o nome, pero da que sufrirán moito nos vindeiros anos", escribiu Keynes sobre o auxe das máquinas que aforran man de obra. A súa predición non se cumpriu no tempo e coa intensidade que el se imaxinou, pero as cousas son moi distintas hoxe día. De feito, estímase que en 2025 o traballo humano e o automatizado equipararanse en cantidade. Para entón, a metade do que podemos considerar traballo será realizado por máquinas. Na maior parte dos casos coexistindo persoas e máquinas, pero tamén, e cada vez máis, substituíndonos estas.

Por outra banda, entre as habilidades e competencias que adquirirán unha maior relevancia en 2025 inclúense a análise e pensamento críticos, a resolución de problemas complexos, a autonomía, a aprendizaxe activa, a resiliencia, a tolerancia a #o #estrés e a flexibilidade. Pode parecernos paradoxal que nun mundo con tanta tecnoloxía e con máquinas facendo o traballo que ata hai pouco faciamos nós, sexan este tipo de competencias, tan humanas, as que máis se demandan e, previsiblemente, máis se demandarán no futuro. Pois precisamente por iso, porque o outro, e cada vez en maior medida, xa o farán as máquinas.

A IA non nos fará prescindibles ás persoas, polo menos non por décadas, pero mudará radicalmente o noso papel. Ninguén está a salvo. Por exemplo, adoita dicirse que os que somos profesores temos un desempeño a proba de automatizacións. Non é certo. Como puxeron de manifesto os estadounidenses Daron Acemoglu e David Autor, automatízase fundamentalmente o que se fai repetidamente e dun modo moi pautado, sexan tarefas físicas ou cognitivas. Por exemplo, é moito máis fácil automatizar o labor dun docente que se limita a presentar contidos e avaliar o progreso dos seus estudantes con exames tipo test, que o traballo que realiza un xardineiro no coidado dos parques municipais.

Neste sentido, e de modo xeneralizado, paso a enunciar o principio de substitución progresiva polo progreso constante: Toda intelixencia [natural] que realiza unha actividade susceptible de ser mellorada ao realizarse por outra intelixencia [artificial], é susceptible de ser substituída por esta.

A cuarta descontinuidade é posible, aínda que estea moi lonxe

Bruce Mazlish foi un prestixioso profesor do departamento de Historia do MIT. Un dos seus libros máis coñecidos titúlase: "A cuarta descontinuidade" e nel fai referencia a unha clasificación feita por Sigmund Freud a principios do pasado século. O pai da psicanálise chama descontinuidades ás grandes decepcións que como humanos fomos vivindo ao fío do avance no coñecemento do mundo e do noso lugar nel. Copérnico descubriu que a Terra non era o centro do universo. Hoxe sabemos que o noso planeta nin sequera é singular no mesmo. Darwin desmontou a idea de que sexamos o resultado da creación directa dun deus. Aínda que moita xente crea e defenda o denominado "deseño intelixente", a ciencia demostrou que procedemos da evolución biolóxica a partir de devanceiros comúns con outras especies que xa pouco teñen que ver connosco, salvo un mesmo tronco na historia da vida. Freud situábase como protagonista da terceira descontinuidade, a través precisamente da psicanálise. Non somos conscientes da maioría das cousas que ocorren no noso cerebro, senón que a consciencia apenas aflora unha parte pequena do que nos goberna e co que tentamos gobernar ou polo menos relacionarnos co mundo que nos rodea.

Mazlish consideraba que a cuarta descontinuidade é aquela na que convivimos coas máquinas, que estenden as nosas capacidades, déixannos atrás naquilo para o que a evolución non nos fixo especialmente competentes, como o cálculo matemático e o razoamento lóxico, e achéganse cada vez máis ás capacidades máis humanas, como o razoamento de sentido común ou a creatividade. As máquinas van ocupando o espazo reservado á intelixencia humana e poden chegar a ser o noso *álter ego* e ata unha alternativa a nós mesmos. No límite, isto suporía a perda da supremacía da intelixencia humana a favor das máquinas. É isto posible?

Sendo certo que a día de hoxe as máquinas, incluso as máis sofisticadas, están lonxe das competencias xerais dun ser humano, non o é menos que os seres vivos chegamos a ser o que somos debido a unha evolución extraordinariamente lenta e sen obxectivos prefixados. Pola contra, nós somos capaces de deseñar con gran rapidez máquinas cada vez máis potentes e capaces e con obxectivos moi concretos. Ademais, neste camiño non é imprescindible que nos inspiremos na bioloxía. Igual que non é necesario simular o voo dos paxaros para sucar os ceos, tratar de imitar un cerebro humano pode non ser a mellor vía para obter unha "superintelixencia" artificial. Por outra banda, se chegamos a esa "superintelixencia", as máquinas xa nos superarían á carreira, porque terían unha capacidade exponencialmente crecente para deseñar a súa vez outras máquinas aínda máis intelixentes. Isto suporía a plena consecución da cuarta descontinuidade enunciada por Mazlish; a primeira que vai máis aló do descubrimento do que somos e do que existe e xorde do que somos capaces de crear artificialmente.

Sobre a viabilidade ou non deste "último fito humano" hai argumentos de todo tipo. O escepticismo abunda máis en persoas alleas ao campo da IA, aínda que sexan moi notables nas súas disciplinas respectivas. É o caso de Noam Chomsky, por exemplo, que considera que esta "singularidade tecnolóxica" é pouco menos que ciencia ficción. Pola contra Stephen Hawking deuno por seguro, aínda que, curándose en saúde, dixo que podemos estar a un cento de anos de logralo. Os que investigamos neste apaixonante campo da intelixencia artificial somos moito máis optimistas que quen traballa en ámbitos das ciencias biomédicas ou na psicoloxía. Quizais por pensar estes que a forma de lograr máquinas cunha intelixencia equivalente á humana será imitándonos a nós, aínda que isto

non sexa nin moito menos necesario. Insisto no exemplo sobre como voamos as persoas fronte ao voar dos insectos ou dos paxaros. Sexa como for, e mentres buscamos logros maiores, a IA é cada vez máis natural nas súas capacidades e na normalidade coa que vai penetrando nas nosas vidas.

En todo caso, de ser posible e de ser logrado, o último invento da humanidade podería ser a superintelixencia artificial. Non é algo que se albisque no horizonte, como xa dixen, pero tampouco que podamos descartar sen máis por pensar que, de ser posible, resultará extraordinariamente complexo logralo. As frases do tipo: "nunca unha máquina será capaz de...", soan un tanto ocas a medida que imos sabendo máis e máis da máquina que así pensa, o noso cerebro, e as intelixencias artificiais son máis e máis competentes. Se algún día as máquinas logran aprender coa flexibilidade coa que nós o facemos, pero moito máis e máis rápido, e deseñar por si mesmas máquinas aínda mellores, chegaría o momento do noso último invento ou polo menos ese será o invento definitivo. Oxalá que para o noso ben. Iso si, cando Mariano Rajoy dixo: "Temos que fabricar máquinas que nos permitan seguir fabricando máquinas, porque o que non vai facer nunca a máquina é fabricar máquinas", xa nos advertiu de que isto é imposible. Ou quixo dicir todo o contrario?

Conservación da intelixencia

Para concluír, quero falar da, aparentemente polo menos, crecente incapacidade e impericia humana para levar a cabo segundo que tarefas. Algo que, de ser así, probablemente ten que ver co uso intensivo, cando non abusivo, da tecnoloxía. De feito, existe un termo acuñado polo sociólogo Harry Braverman, para reflectir o impacto que provoca a automatización do traballo nunha menor cualificación da man de obra, Trátase de "deskilling", que poderiamos traducir, aínda que dun modo un pouco forzado, como descualificación. Son múltiples os exemplos que podemos dar desta descualificación progresiva, e máis os que irán aparecendo nos próximos anos. Segundo parece, a impericia dos pilotos tivo moito que ver no accidente aéreo de 2009 dun voo de Air France, que voaba de París a Rio de Janeiro, a pesar dos miles de hora de voo de experiencia que estes tiñan. Algo semellante pode ocorrer coa condución de vehículos, a medida que a autonomía destes creza; a redución das habilidades diagnósticas humanas baseadas en analíticas, imaxes médicas ou sinais fisiolóxicos, na medida en que cada vez máis están soportadas ou apoiadas polas máquinas; ou o incremento nos erros ortográficos e a redución nas habilidades de comunicación escrita, debido ao uso cada vez máis común de correctores e outro tipo de tecnoloxías lingüísticas. Estamos a experimentar unha involución humana?

No ámbito profesional esta descualificación podería levarnos de traballadores de pescozo azul e pescozo branco a aqueles sen pescozo, en camiseta, desempregados e que perderían ou nin sequera adquirirían habilidades humanas importantes –habilidades matemáticas e de razoamento espacial, mesmo elementais, competencia en idiomas, saber escribir correctamente, tanto ortograficamente como na calidade literaria do redactado.

Pero pode ser moito peor. Dinos Francisco Moura que hai voces prestixiosas que alertan do dano cerebral que poden producir as TIC nos nenos. Por exemplo, navegar por Internet require dun foco de atención moi curto e cambiante, o que pode ir en detrimento de a

atención executiva. Son moitas as consecuencias que poden derivarse diso, como a diminución da empatía, da atención executiva e do autocontrol, ou o aumento da impulsividade, a diminución na capacidade de tomar decisións e, por suposto, provocar adiccións.

Resulta paradoxal pensar que o incremento da intelixencia das máquinas poida supoñer unha diminución na intelixencia humana. É coma se a intelixencia no mundo sometésese ás regras dos xogos de suma cero. É dicir, coma se a cantidade de intelixencia do mundo mantivésese constante, de modo que a que gañan as máquinas perdémola nós. De feito, hai un chiste que circula moito na internet no que unha persoa dille a outra que non está preocupada polo aumento da IA senón pola diminución da intelixencia humana.

Sería moi grave que a cociente entre IA e intelixencia humana estea a aumentar non só por aumentar o numerador, o cal é positivo de entrada, senón porque diminúe o denominador, o que non se se é certo, pero é posible. É máis, a día de hoxe formulo só como unha conxectura a constancia na suma das intelixencias artificial e humana, pero estou a recompilar evidencias que poderían facer que a conxectura pase a ser un principio. Oxalá non sexa así.

Santiago de Compostela, 10 de xaneiro de 2023