

TERMINÀLIA parla amb...

Ramon López de Mántaras i Badia, director de l'Institut de Recerca en Intel·ligència Artificial



Ramon López de Mántaras i Badia Sant Vicenç de Castellet, 1952

Ramon López de Mántaras és doctor en Física per la Universitat Paul Sabatier de Tolosa i doctor en Informàtica per la Universitat Politècnica de Catalunya. Va fer estudis a la Universitat de Mondragón i a la Universitat de Califòrnia. És professor del Consell Superior d'Investigacions Científiques (CSIC), on dirigeix l'Institut de Recerca en Intel·ligència Artificial, ubicat al campus de la Universitat Autònoma de Barcelona.

És un dels pioners de la intel·ligència artificial a Catalunya i a Espanya. Ha rebut diversos premis: el Premi Ciutat de Barcelona d'Investigació (1982), el Premi Robert S. Engelmores de l'Associació Americana d'Intel·ligència Artificial (2011), el Premio Nacional de Informàtica de la Societat Científica Informàtica d'Espanya (2012), l'EurAI Distinguished Service Award de l'Associació Europea d'Intel·ligència Artificial (2016), el

Donald E. Walker Distinguished Service Award de la International Joint Conferences on Artificial Intelligence (2017) i el Premio Nacional de Investigación Julio Rey Pastor en Matemáticas y Tecnologías de la Información y las Comunicaciones (2018). L'any 2016 va ingressar com a membre numerari a l'Institut d'Estudis Catalans.

Els productes d'intel·ligència artificial (IA) com el ChatGPT s'han fet molt populars darrerament. La possibilitat d'interactuar-hi amb llenguatge natural, com ho faries xerrant amb una altra persona, probablement ha contribuït a la seva ràpida difusió. De fet, el ChatGPT es defineix com un *large language model* (LLM). També n'hi ha que el defineixen com un *large multimodal model* (LMM). Què és un *language model* (LM) i un *large language model*, i quina és la diferència entre aquests models i els *large multimodal models*?

Un LM és un model computacional de llenguatge que es basa en el càlcul de la probabilitat de seqüències de paraules. És a dir, calcula la probabilitat que unes paraules vagin a continuació d'unes altres en un idioma determinat. Un LM s'entrena només a partir de text. En canvi, un LLM és un model computacio-

nal (basat en xarxes neuronals artificials profundes) que s'ha entrenat amb un corpus enorme i multimodal (text, imatges, veu), i això fa que l'avaluació probabilística de quines paraules segueixen unes altres paraules sigui més acurada. A més de text, els LLM poden generar imatges basant-se en descripcions textuales o imatges introduïdes per l'usuari, i també poden generar textos molt més llargs que els LM i en diversos idiomes.

Quines aplicacions tenen els LLM en la vida quotidiana?

Poden tenir diverses aplicacions. Les més importants són: bots per xatejar, traducció automàtica, generació de textos (notícies, informes financers, redacció de contractes, respostes a correus electrònics...), transcripció de veu a text, assistents virtuals, resum d'articles i documents, recomanació de productes

(música, pel·lícules). Tot i que s'estan aplicant en sectors com ara la medicina, cal anar molt amb compte quan es tracta d'aplicacions en què la precisió de la informació és crítica. Per exemple, en un estudi titulat «Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine», publicat al *New England Journal of Medicine* el 30 de març de 2023, els metges van avaluar la fiabilitat del GPT-4 formulant la mateixa pregunta diverses vegades durant diversos dies. Utilitzant mètriques comunes per avaluar la similitud de les respostes generades, es va trobar una similitud molt baixa i una gran variabilitat entre les respostes a la mateixa pregunta. Globalment, només el 41% de les respostes de GPT-4 coincidien amb respostes reconegudes com a correctes en medicina.

Els LLM han tingut molta rellevància en aplicacions lingüístiques com la traducció automàtica neuronal (TAN)



Font: arxíu personal

Les tecnologies com la traducció automàtica o la generació creativa de text son molt lluny de tocar sostre

o la generació creativa de text. Com a tecnologia lingüística, han tocat sostre? Quina aplicació, present o futura, tenen?

No! Som molt lluny de tocar sostre. De fet, tot i l'aparent capacitat de comprensió del llenguatge, de fet no comprenen res, tret que acceptem una definició molt restrictiva i superficial del que vol dir *comprendre*. Si *comprendre* es limita a generar

cadena de text que els humans trobem coherents, aleshores podríem dir que comprenen el llenguatge. Però, en la meua opinió, i d'acord amb el que deia Saussure, el llenguatge està format pel significat i el significat, i els LLM només tracten (s'entrenen) amb el significat i, per tant, no hi ha cap possibilitat que aprenguin el component del significat, ni semànticament ni pragmàticament ni culturalment parlant. Estic força d'acord amb l'experta en lingüística computacional Emily Bender quan diu que els LLM són «lloros estocàstics». En un futur, l'ús d'aquests models en IA corpòries (és a dir, en robots físics) pot ser una via per començar a superar la manca de coneixements de sentit comú (coneixements sobre el món), ja que una IA amb cos permetria experimentar amb el món i

així començar a aprendre a relacionar el significat amb el significat. És un problema obert des que existeix la IA (a mitjans dels anys cinquanta) i es coneix com el «problema de la base simbòlica» (*symbol grounding problema*, en anglès).

Com a lingüistes, sovint ens costa assimilar la idea que un LM i un LLM calculen o processen llengua. Què té d'especial la IA que proporciona o analitza textos en relació amb la IA que proporciona respostes en altres formats (imatges, dades numèriques, etc.)?

La IA és molt més que els LLM i l'aprenentatge profund (DL, *deep learning*). Hi ha molts algoritmes que tenen poc o res en comú amb els dels LLM. Per exemple, en visió per computador s'usen xarxes convolucionals per identificar objectes en imatges que són diferents dels que s'usen en els LLM. En jocs de tauler com ara el go o els escacs s'usen combinacions d'algoritmes basats en xarxes neuronals profundes i algoritmes d'aprenentatge per reforç que són també molt diferents. En planificació i raonament s'usen tècniques

Els LLM només s'entrenen amb els significants i no hi ha cap possibilitat que aprenguin el component del significat, ni semànticament ni pragmàticament ni culturalment parlant

d'IA, basades en la lògica matemàtica, que també són completament diferents. En sistemes multiagent, en què els problemes es resolen mitjançant la col·laboració entre diversos agents intel·ligents, les tècniques d'IA també són molt diferents. En definitiva, en IA tenim un

ventall molt ampli d'algoritmes molt diversos que s'usen, individualment o en combinació, en funció del problema que cal resoldre. Per exemple, el famós sistema AlphaFold, que prediu el plegament de proteïnes, combina trenta-dos algoritmes diferents. No existeix un únic algoritme general (és a dir, una IA general) capaç de dur a terme qualsevol tasca.

Un altre aspecte que ens costa assimilar com a lingüistes és que pràcticament totes les dades són aptes per entrenar una IA. Per exemple, en traducció automàtica estadística (TAE) i en altres sistemes entrenats amb quantitats de dades inferiors a les utilitzades en un LLM, es tenia

Els LLM no tenen una comprensió real de la

llengua i del món com tenim nosaltres, els humans.

Només detecten patrons lingüístics comuns en el corpus d'entrenament

molt en compte la màxima *garbage in, garbage out*, que ve a dir que un sistema entrenat amb dades de poc valor proporcionarà resultats que també seran d'una qualitat baixa. Certament hem superat aquesta limitació?

No, no hem superat aquesta limitació. De fet és molt evident que els LLM continuen generant *garbage* i això és un problema molt seriós, ja que generen respostes esbiaixades pel que fa a aspectes com ara gènere, racisme... Aquestes respostes s'intenten evitar introduint filtres de manera manual, però això no és cap solució real, ja que només són pedaços que no podran solucionar mai un problema que és inherent a l'absoluta manca de comprensió del

llenguatge per part dels LLM. Això, juntament amb el problema dels múltiples errors factuais que generen, que fa que siguin uns grans productors de desinformació, pot causar polarització en la societat i posar en perill fins i tot la democràcia. Com molt bé diu Daniel Dennett, els sistemes com el ChatGPT generen «counterfeit people» i, si falsificar moneda és un crim, falsificar persones hauria de ser-ho també.

I una capacitat més, que sembla contradir alguns dels principis lingüístics (i particularment traductològics) anteriorment inamovibles: la generació de text en una llengua a partir de sistemes entrenats amb textos en altres llengües. Què fa que un sistema de generació de text com un LLM pugui produir text en una llengua que està molt poc representada en el seu corpus d'entrenament?

Els LLM han estat entrenats amb textos en moltes llengües diferents i dissenyats per detectar-hi patrons lingüístics, és a dir, relacions entre paraules, frases i estructures gramaticals. De fet, detectar patrons és el que realment saben fer les xarxes neuronals profundes. Com que aquests patrons són comuns a moltes llengües, els LLM els utilitzen per generar textos en llengües fins i tot poc representades en el conjunt d'entrenament. Ara bé, cal deixar clar que els resultats en llengües poc representades són de qualitat força inferior. Òbviament, en el cas de llengües que, a més d'estar poc representades en els corpus d'entrenament, tenen estructures lingüístiques molt diferents, els resultats seran molt deficients.

Parlem ara dels productes textuals d'aquests sistemes. N'hi ha que s'apropen amb reticències als productes, com el ChatGPT, a causa de l'homogeneïtat del text produït. Per exemple, qualsevol text analític o reflexiu que escrivia el ChatGPT3 acostumava a acabar amb un paràgraf introduït pel connector «Overall, ...». Aquests límits creatius són

insuperables? O arribarà el sistema que sigui capaç d'emular la manera de fer servir la llengua de cada usuari?

Tal com he respost en una de les preguntes anteriors, aquests models no tenen una comprensió real de la llengua i del món com tenim nosaltres, els humans. Només detecten patrons lingüístics comuns en el corpus d'entrenament. Per això, els textos que generen són homogenis. En teoria, es podria millorar el que fan actualment si s'entrenessin amb un corpus suficientment ampli específic per a cada usuari. Òbviament, llevat d'alguns autors, la immensa majoria dels usuaris no han generat un corpus prou ampli de textos

La recerca en terminologia, pot contribuir-hi

millorant, per exemple, la comprensió del

llenguatge especialitzat

per entrenar un LLM. D'altra banda, aquesta personalització, amb els LLM actuals, seria molt costosa tant en temps com en termes econòmics. Caldrien models computacionals diferents dels LLM, que permetessin aprendre la manera de fer servir la llengua de cada usuari partint de corpus relativament petits.

Com a lingüistes, quines eines d'IA poden contribuir a la nostra recerca? I viceversa, quina recerca nostra pot contribuir al desenvolupament de la IA amb aplicacions de llengua o lingüística? En particular, la recerca en terminologia i els productes terminològics tenen cabuda en l'escenari futur de desenvolupament de la IA?

Crec que hi ha treballs que han desenvolupat molts lingüistes computacionals i també experts en *natural language understanding* (NLU) al llarg dels darrers cinquanta anys que



Foto: arau personal

podrien ser molt útils per millorar els models actuals. Per exemple, mitjançant sistemes híbrids que aprofitin els avantatges de les aproximacions antigues, basades en representacions computacionals del coneixement que atacaven els problemes de les dimensions semàntica, pragmàtica i cultural de la llengua, i les aproximacions actuals, basades en la detecció de patrons en dades reals. Aquesta hibridació potser permetria desenvolupar models de llenguatge més «personalitzables» (vegeu la pregunta anterior).

Pel que fa a la recerca en terminologia, pot contribuir-hi millorant, per exemple, la comprensió del llenguatge especialitzat aportant recur-

sos terminològics i diccionaris per entrenar i millorar els algorismes.

I una darrera pregunta, relacionada amb la terminologia. La popularització del ChatGPT ha comportat també un vocabulari associat. Hi ha un terme en particular que s'utilitza molt: *prompt*. Ja es parla de *prompt engineering* i de *ser expert en prompts*. Què és un *prompt*? I com el traduiries al català?

És una entrada en forma de text (frase, paràgraf, pregunta...), més aviat breu, que s'utilitza per iniciar la generació d'un text per part del ChatGPT i també per orientar sobre quin tipus de text volem obtenir i de quin tema. Aquest text pot ser

una continuació de la frase o de les frases d'entrada o una resposta en cas que l'entrada sigui una pregunta. Com més òptim sigui el *prompt* millor serà la qualitat de la resposta generada, per això és important l'habilitat de l'usuari a l'hora de dissenyar el *prompt*, i d'aquí sorgeix això del *prompt engineering* per referir-se a l'habilitat de crear *prompts* efectius.

No tinc clar quina seria la millor traducció. N'he vist diverses opcions en la bibliografia científica, com *entrada*, *instrucció*, *consigna* o *indicació*.

CONSELL DE REDACCIÓ