



1187 Hinterwasser

Quizás sean las ganas de hablar de algo realmente intrascendente, pero el caso es que la presentación en sociedad de un artefacto conocido como ChatGPT ha causado un revuelo inusual en amplios pagos de la comidilla social de este planeta. Es un pretendido nuevo hito de la Inteligencia Artificial que persigue reinar en nuestro incipiente siglo, pero quizás las cosas no estén tan claras como se están tratando en los medios masivos y generalistas muchas veces plagados de nuevos tecno-evangelistas que prometen “el oro y el moro”. No está de más echarle un vistazo en esta columna a lo que ese artefacto, el ChatGPT y sus análogos, pueden realmente aportar y significar para nuestra realidad presente y futura.

“1187 Hinterwasser. Ese es mi hotel. ¿Qué? Donde yo vivo. ¿Un sitio bonito? Sí, eso creo. ¿Esto es parte del test? No, era sólo por charlar. Esta usted en un desierto, caminando por la arena, cuando... ¿Eso ya es el test? Sí. Está usted en un desierto, caminando por la arena, cuando, de repente... el galápagos yace sobre su espalda con el estómago cociéndose al sol y moviendo las patas para darse la vuelta, pero sin su ayuda no puede. Y usted no le ayuda. ¿Qué quiere decir que no le ayudo? Quiere decir que no le ayuda. ¿Por qué es así León?...

Solo son preguntas. Es un test hecho para provocar una respuesta emocional. [pausa] ¿Quiere que sigamos? Describame, con palabras sencillas, sólo las cosas buenas que le vienen a la mente... acerca de su madre. ¿Mi madre? Sí. Le voy a hablar de mi madre... León dispara a su interrogador repetidas veces y lo mata al instante”.

Esta es una de las escenas iniciales de la magnífica película de culto “Blade Runner”¹ que ya en 1982 planteaba un montón de preguntas y cuestiones que las nuevas generaciones creen que son de ahora porque las relacionan con el actual e interesado reverdecir de la publicidad y financiación de la denominada Inteligencia Artificial.

Para aquellos que puedan no saberlo ya, el motivo central de la historia y la existencia del *Blade Runner* protagonista de la historia es una fuga de las “colonias del mundo exterior” de seis “replicantes”, tres varones y tres hembras. En ella asesinaron a 23 personas y asaltaron una lanzadera que una patrulla aérea divisó lejos de la costa (de San Francisco en 2019); sin ningún triplante y ni rastro de ellos. Los replicantes intentaron entrar en la Tyrell Corporation en busca de algunas respuestas existenciales y

un nuevo futuro, pero uno de ellos se abrasó al atravesar “un campo electromagnético” mientras la policía perdía la pista de los cinco restantes. A la vista de que hubieran podido infiltrarse como empleados, la policía mandó a un agente, a un *blade runner*, para que efectuara un test Voight Kampff² a los nuevos trabajadores y, sin duda, descubrió a uno de ellos.

Los replicantes fueron diseñados por la Tyrell Corporation como copias de seres humanos en todos los sentidos, excepto en

vimiento de 2022, de uno de sus productos que es el ChatGPT⁵ (Generative Pre-trained Transformer) que es una chatbot⁶ construido sobre una familia de modelos de lenguajes de Lenguaje Natural de OpenAI (el GPT-3 para ser más precisos), operando juntos a sistemas de aprendizaje automático supervisado⁷ y reforzado⁸. OpenAI ha estimado en 29 billones de dólares el valor de su juguete⁹.

En su programa de entrenamiento, seres humanos juegan el papel de usuario



Muchas veces se pretende que la IA aprenda cosas que realmente no sabemos enseñarle y recurrimos a grandes volúmenes de datos extraídos de la realidad cotidiana; lo más lejos que puede ir el ente entrenado así es al valor promedio, al sentido común, a la mediocridad de lo real. Si se aprende de los datos, lo único que se destilaría sería la mediocridad humana (in sensu estadístico).

sus emociones. Pero, los diseñadores creían que, al cabo de unos años, podían desarrollar sus propias emociones; odio, amor, miedo, cólera, envidia... y por eso les dotaron con un mecanismo de seguridad. ¿Cuál es? Sólo cuatro años de vida.

El revuelo mediático actual lo ha causado un lanzamiento de la compañía OpenAI³ del controvertido y despótico Elon Musk⁴. En concreto, el lanzamiento, el 30 de no-

y, a su vez, de Inteligencia Artificial (ellos se lo guisan y ellos se lo comen) generando así un conjunto de respuestas para cualquier pregunta o entrada que haya sido planteada. Posteriormente, se evalúan cada una de las respuestas de mejor a peor en cuanto a su calidad, y las calificaciones obtenidas en esta evaluación se utilizan para “recompensar” (incentivar) al modelo. Esto se repite *sine die* con la

¹ Ver https://en.wikipedia.org/wiki/Blade_Runner

² “¿Esto va a ser un test de empatía? Dilatación capilar por las así llamadas respuestas ruborizantes. Fluctuación de la pupila. Dilatación involuntaria del iris, ... Nosotros lo llamamos Voight Kampff para abreviar.”

³ Ver <https://en.wikipedia.org/wiki/OpenAI>

⁴ Ver https://en.wikipedia.org/wiki/Elon_Musk

⁵ Ver <https://en.wikipedia.org/wiki/ChatGPT>

⁶ Ver <https://en.wikipedia.org/wiki/Chatbot>

⁷ Ver https://en.wikipedia.org/wiki/Supervised_learning

⁸ Ver https://en.wikipedia.org/wiki/Reinforcement_learning

⁹ Ver <https://www.eweek.com/big-data-and-analytics/chatgpt/>

esperanza de que la IA aprenda algo.

El algoritmo de aprendizaje reforzado de OpenAI se llama Proximal Policy Optimization¹⁰, y se basa en una técnica similar al Stochastic Gradient Descent¹¹, lo que le da mucha velocidad de respuesta con un coste computacional menor de lo que es habitual en las redes de Procesamiento de Lenguajes Naturales (NLP)¹².

OpenAI utilizó la infraestructura de supercomputación de Azure en el proceso de entrenamiento de ChatGPT que se completó a finales de 2021. ChatGPT utiliza las entradas previas de la misma sesión para generar respuestas mucho más precisas y contextualmente relevantes. Para evitar males mayores y grandes planchazos, ChatGPT cuenta a su salida con un filtro moderador que intenta eliminar el racismo, el sexismo, los sesgos de todo tipo o cualquier otro contenido que pueda ser ilegal u ofensivo.

Limitaciones de la plataforma

La plataforma ChatGPT actual tiene algunas limitaciones. Algunas veces da respuestas sin sentido, tiene cierta tendencia a ser prolija en sus respuestas, y tiene una clara incapacidad de plantear preguntas aclaratorias que le permitan responder preguntas del usuario que sean ambiguas. En algunos casos, cambiar una o dos palabras puede alterar dramáticamente la respuesta con la que nos agasaja ChatGPT.

El “*machine learning*”¹³ subyacente en las IAs que alcanzan audiencia de masas me recuerda a otro sueño de la humanidad en el que juega a ser dioses creadores, y es el de los Golems. El ejemplo más conocido de este ensueño es el de los rabinos de Praga (Rabbi Loew¹⁴ en particular) que soñaban con crear Golems¹⁵ a partir de la arcilla y dotarles de vida de modo que éstos hicieran los encargos que ellos no podían o no querían hacer¹⁶. El Golem es fuerte, pero no inteligente; y si se le ordena llevar a cabo una tarea, la llevará a cabo de un modo sistemático, lento y ejecutando las instrucciones de un modo literal, sin cuestionamiento ninguno (un verdadero ideal para muchos Departamentos de Recursos Humanos).

Tanto en las versiones históricas como en las actuales, el sueño que subyace en

este tipo de escenarios es la posibilidad que pueda tener o no el ser humano de generar *ex nihilo* otros seres capaces y/o inteligentes; tan capaces como él o incluso superiores. En el fondo está la soberbia atávica de jugar a ser dioses, de crear seres inteligentes, olvidándose (¿no sé por qué?) del proceso estándar de concepción que compartimos con todos los mamíferos de la tierra.

La pregunta es ¿por qué queremos crear seres inteligentes? ¿Para qué tanto interés en las mal llamadas Inteligencias Artificiales (que no son inteligentes, pero sí artificiales)? Pensar que es simplemente por atender al reto, es algo bastante *naïve* ya que el capital “*no da nunca puntada sin hilo*” y algo espera obtener a cambio. Pensándolo bien, ChatGPT no es más que la continuación de los intentos clásicos para desarrollar Modelos de Lenguaje Natural que permitiesen canalizar “el conocimiento” hacia y entre sistemas “inteligentes”. Dicho de otro modo, ChatGPT podría verse como un pariente avanzado de Siri¹⁷, Alexa¹⁸ o Cortana¹⁹. La idea es que puedan interactuar con los humanos a través del mecanismo más antiguo y característico que tenemos; el del habla, tanto si es locutada como si es escrita.

Los ChatBots no son nada nuevo, ya tuvieron su amanecer hace tiempo cuando

ChatGPT viene a ser una nueva ofensiva de dichos comerciantes en busca de lugar para sus mercancías. Es cierto que los tiempos que corren están abriendo frentes que pueden llegar a ser importantes. Un ejemplo de ello sería el uso de los ChatBots para hacer compañía a personas solitarias por edad o por falta de socialización (que son muchas en ambos casos).

Por ejemplo, Replika²⁰ es uno de los mejores chatbot de 2022 y se ha centrado en proveer de ayuda (conversación) a personas que realmente necesitan apoyo con sus problemas de soledad y aislamiento. En una sociedad en la que hablar, comunicarse con los demás, se ha convertido en algo difícil (especialmente con la edad), los autómatas con capacidad de dialogo pueden resultar atractivos para las personas que necesitaban socializar. Si los hay que hablan con sus plantas²¹, y hablan con sus mascotas, ¿Por qué no hacerlo con un ChatBot?

Propuestas y temores de la presentación en sociedad de ChatGPT

Con la presentación en sociedad de ChatGPT han sido numerosísimas las propuestas y temores que ha despertado. En su afán marquetiniano de “*Lo importante es que hablen de ti, aunque sea bien*” (Sal-



No hay que olvidar que la información NO ES inteligencia; por lo menos no es la inteligencia de la que se habla cuando nos referimos a seres humanos. Y que algo parezca humano NO lo hace humano.

las empresas creyeron que podían sustituir sus “Call Centers” de humanos mal pagados por inteligencias artificiales que estarían allí, a todas horas, pasase lo que pasase, interesados sólo en atender al usuario/cliente en lo que tuvieran a bien plantearles. Si el hartazón con los Centros de Llamadas era significativo entonces, el enfrentarse inútilmente a un tenaz artefacto informático que sólo conoce cuatro reglas (lógicas) de las cuales no se va a desviar, ha causado en los usuarios/clientes un rechazo frontal que ha enfriado bastante aquellas alegrías iniciales de los comerciantes de IAs.

vador Dalí), algunos contertulios/cuñaos de amplio espectro, que trabajan en los medios de información de masas, se han preguntado si ChatGPT puede llegar a ser la muerte y el final del periodismo tal y como lo conocemos, si puede terminar con la radio desarrollada por humanos, y si terminará copando la comunicación de masas en general. ¿Por qué seguir a un *influencer* si podemos seguir a una ChatBot?

Otros incluso han dicho que ¿para qué queremos maestros en las escuelas si podemos tener chatbots que puedan educar/formar a nuestros hijos y, además, puedan “correr” en sus teléfonos móviles? Los hay que, en su afán indagador, en su curiosidad, le han pedido a ChatGPT que componga un RAP o un cuento con un determinado tema, y el autómata lo ha compuesto. Otros le han pedido que escriba un programa de ordenador y algo les ha programado (y quizás puede sea mejor que algunos productos de las Software Factories). No se ha librado de este furor pedirle al artefacto que juegue distintos papeles en escenarios de ciberseguridad. Incluso hay algunos que se preguntan sobre su posible impacto en mundo

¹⁰ Ver https://en.wikipedia.org/wiki/Proximal_Policy_Optimization

¹¹ Ver https://en.wikipedia.org/wiki/Stochastic_gradient_descent

¹² Ver https://en.wikipedia.org/wiki/Language_model

¹³ Ver https://en.wikipedia.org/wiki/Machine_learning

¹⁴ Ver https://en.wikipedia.org/wiki/Judah_Loew_ben_Bezalel

¹⁵ Ver <https://en.wikipedia.org/wiki/Golem>

¹⁶ Ver https://en.wikipedia.org/wiki/Golem#The_classic_narrative:_The_Golem_of_Prague

¹⁷ Ver <https://en.wikipedia.org/wiki/Siri>

¹⁸ Ver https://en.wikipedia.org/wiki/Amazon_Alexa

¹⁹ Ver [https://en.wikipedia.org/wiki/Cortana_\(virtual_assistant\)](https://en.wikipedia.org/wiki/Cortana_(virtual_assistant))

²⁰ Ver <https://replika.ai/>

²¹ Ver <https://jardinessinfronteras.com/2020/08/24/la-terapia-horticola-y-los-jardines-terapeuticos/>

de la abogacía y los pleitos²². Como toda (presunta) nueva tecnología, son muchos los miedos y rechazos que levanta.

Errado el tiro

Sin embargo, creo que esos miedos tienen errado el tiro. Los promotores de los ChatBots, al modo de Golems medievales, lo que quieren es poder fabricar trabajadores automáticos que sustituyan a trabajadores humanos, que son mucho más “caros” (a ojos del que les mal-paga), y que tienen derechos sindicales y laborales, además de otros derechos como seres inteligentes y sintientes (Derechos Humanos).

La idea de la IA comercial es llegar, en el universo robótico, a condiciones de trabajo (24/7/365) sin más pago/coste que la inversión y su sustento energético, para alcanzar eficiencias no imaginadas. Estas condiciones son ilegales en los humanos desde hace casi un siglo pero, claro está, los robots no son seres humanos.

¿Para qué queremos poder fabricar equivalentes funcionales de ciudadanos mediocres? Para hacerles trabajar sin ningún tipo de derechos. ¿Tendrán derechos civiles y sindicales las IAs o cualquier otro sistema robótico? Lo dudo mucho. Lo bueno de esta distopía en ciernes es que no va a ser posible, ya que la IA comercial tiene los pies de barro.

Centrándonos en el escenario de todo lo basado en *machine learning* y *data mining*, es conveniente recordar que sólo se puede enseñar lo que uno realmente sabe. La calidad de las fuentes de información es un factor limitante superior en todo proceso de enseñanza, y es del todo insalvable. Por otra parte, las técnicas de entrenamiento de estas Inteligencias Artificiales son netamente paulovianas²³; los aciertos son premiados y los errores penalizados²⁴. Sin embargo, no hay que olvidar que la **información NO ES inteligencia**; por lo menos no es la inteligencia de la que se habla cuando nos referimos a seres humanos.

Si a este enfoque le sumamos que algunas/muchas veces se pretende que la IA aprenda cosas que realmente no sabemos enseñarle, y que recurrimos a los grandes volúmenes de datos extraídos de la realidad cotidiana (*data mining*), los más lejos que puede ir el ente entrenado así es al valor promedio, al sentido común, a la mediocridad de lo real. Si se aprende de los datos, lo único que se destilaría sería la mediocridad humana (*in sensu estadístico*).

En el mejor de los casos, las IAs estadísticamente entrenadas podrían llegar a ser tan inteligentes como el común de los mortales que, en realidad, no es suficientemente inteligente como para resolver problemas serios. Que algo parezca humano²⁵ NO lo hace humano.

La superioridad de la genialidad humana, aunque sólo sea a fuerza de tener miles de millones de posibles candidatos sobre el planeta, está muy por encima de lo que podría entrenarse a cualquier Golem, clásico o del Siglo XXI. ¿Puede una IA ser disléxica? ¿Puede entrenarse la creatividad o la genialidad? Me temo que la respuesta es no en todos los casos.

Los algoritmos de *machine learning* no logran superar la rigidez intrínseca que imponen los sistemas computacionales secuenciales del universo digital actual. Evolutivamente no se ha conseguido una materia tan plástica para el aprendizaje como lo son los cerebros orgánicos en general, y el de los primates en particular. ¿Por qué gastar tantos recursos y energía en enseñar a las muy limitadas IAs cualquier cosa y no dedicar esos recursos a formar seres humanos?



La codicia económica y los dioses del máximo beneficio del capital, no justifican la creación de cualquier tipo de “pellejudos”, androides o no, chatbots o asistentes de ayuda en la toma de decisiones, a los que despojar de los derechos de la humanización.

¿Por qué debemos invertir en IAs en lugar de hacerlo en los sistemas de enseñanza clásicos cuyo material y objetivo son las mentes y cerebros humanos? La respuesta es, probablemente, porque los humanos tienen derechos y, hoy por hoy, no así las máquinas.

En la película A.I. Raising²⁶ (2018) del director Lazar Bodroža, basada en una historia corta de Zoran Nešković, su protagonista Milutin, un astronauta bien entrenado y experto en lo suyo, no tarda ni unos pocos días en violar inmisericordemente a Nimani, un androide con una (magnífica) apariencia femenina diseñada para responder a los deseos del humano y para monitorizar su trabajo en la nave. ¿Es que una violación deja de ser violación por realizarla contra un androide? La película ahonda en los confines de la naturaleza humana, y la presencia de androides y sistemas de inteligencia artificial es meramente una excusa estética y

un recurso discursivo que ayuda a resaltar lo que realmente es el centro de todo: la naturaleza humana.

Deshumanizar a los humanos y humanizar a lo inhumano

Hay que tener cuidado con deshumanizar a los humanos y con humanizar a lo inhumano; cada ente tiene su materia esencial y sirve para lo que ya se ha visto (históricamente) que sirve. Siempre que se han saltado esos límites, siempre que unos han intentado sustituir a los otros, las cosas han terminado siendo un desastre. La codicia económica y los dioses del máximo beneficio del capital, no justifican la creación de cualquier tipo de “pellejudos”²⁷, androides o no, chatbots o asistentes de ayuda en la toma de decisiones, a los que despojar de los derechos de la humanización.

No caigamos en la distopía de la Tyrell Corporation: “*El comercio es nuestro objetivo aquí, en la Tyrell. Y nuestro lema ‘más humanos que los humanos’.* Rachael es un experimento, nada más. Empezamos a per-

cibir en ellos extrañas obsesiones. Después de todo son inexpertos emocionalmente. Con unos años para almacenar las experiencias que usted y yo damos por supuesto. Si les obsequiamos con un pasado, creamos un apoyo para sus emociones y, consecuentemente, podemos controlarlos mejor. ¡Recuerdos! Usted habla de recuerdos”.

Yo prefiero pensar como en el epílogo de la novela The Time Machine²⁸ de H.G. Wells: “Si esto es así, no nos queda sino vivir como si no lo fuera. Pero, para mí, el porvenir aparece aún oscuro y vacío; es una gran ignorancia, iluminada en algunos sitios casuales por el recuerdo de su relato. Y tengo, para consuelo mío, dos extrañas flores blancas –encogidas ahora, ennegrecidas, aplastadas y frágiles– para atestiguar que aun cuando la inteligencia y la fuerza habían desaparecido, la gratitud y una mutua ternura aún se alojaban en el corazón del hombre.”... y no en la IA al uso. ■

²² Ver <https://elderecho.com/reflexiones-sobre-el-impacto-de-chatgpt-en-la-profesion-legal>

²³ Ver https://en.wikipedia.org/wiki/Classical_conditioning

²⁴ Ver <https://www.machinelearningpro.org/what-is-machine-learning-model-training/>

²⁵ Ver https://en.wikipedia.org/wiki/Turing_test

²⁶ Ver https://en.wikipedia.org/wiki/A.I._Rising

²⁷ Blade Runner film (1982): “Pellejudos. Así era como (el capitán) Bryant llamaba a los replicantes. En los libros de historia él es el tipo de policía que solía llamar chimpancés a los negros.”

²⁸ Ver https://en.wikipedia.org/wiki/The_Time_Machine

JORGE DÁVILA

Consultor independiente

Director

Laboratorio de Criptografía

LSIS – Facultad de Informática – UPM

jdavila@fi.upm.es