Estudio de la influencia de la inteligencia artificial GPT-2 y su generación de textos, mediante el procesamiento de lenguaje natural, en la valoración de estos como veraces por parte de los estudiantes de 11vo grado del colegio San Agustín de Lima, Perú.

¿En qué medida la inteligencia artificial GPT-2 y su generación de textos, mediante el procesamiento de lenguaje natural, influyen en la valoración de estos como veraces por parte de los estudiantes de 11vo grado del colegio San Agustín de Lima, Perú?

Tecnología de la Información para una Sociedad Global

Nivel Medio

Código personal: jnk726

Número de palabras: 3978

Esta Monografía debe referenciarse de la siguiente manera:

Flores, A. (2021). *Estudio de la influencia de la inteligencia artificial GPT-2 y su generación de textos, mediante el procesamiento de lenguaje natural, en la valoración de estos como veraces por parte de los estudiantes de 11vo grado del colegio San Agustín de Lima, Perú.* [Trabajo de investigación. Monografía, Centro Educativo Particular San Agustín] Perú.

ÍNDICE

INTRODUCCIÓN

¿Unicornios en los Andes? Un día, mientras navegaba por Internet, encontré esta insólita noticia. Un grupo de investigadores habían descubierto unicornios hablantes. Increíblemente, el texto fue redactado por una inteligencia artificial. Probablemente, nunca existan los unicornios; no obstante, la empresa OpenAI ha creado algo más interesante, a GPT-2.

Ante la situación, me invadía un sentimiento de alegría y preocupación. Por un lado, GPT-2 era increíble, por el otro, temía la viralización de textos falsos y la credibilidad de la población. Debido a ello, surge la pregunta de investigación: ¿En qué medida la inteligencia artificial GPT-2 y su generación de textos, mediante el procesamiento de lenguaje natural, influyen en la valoración de estos como veraces por parte de los estudiantes de 11vo grado del colegio San Agustín de Lima, Perú?

En la actualidad, la información falsa es escrita por humanos. Debido a ello, por fortuna, su producción se encuentra limitada. Sin embargo, la situación podría empeorar, el campo de procesamiento de lenguaje natural está creciendo a pasos agigantados, volviendo factible la automatización de esta maliciosa actividad; es aquí donde radica la importancia del presente trabajo.

Considero que el tema es significativo por la presencia de personas malintencionadas en difundir información falsa, siendo altamente probable que cuando el procesamiento de lenguaje natural mejore, no duden en explotarlo. Finalmente, la monografía trasciende al evaluar la situación actual, estableciendo un punto de partida para afrontar la futura amenaza.

METODOLOGÍA

A fin de brindar un marco teórico referente a la pregunta de investigación, el primer capítulo profundizará en su variable dependiente e independiente. Sobre la dependiente, se abordará contenido acerca del concepto información e información falsa, centrándose en GPT-2 y el impacto de esta cuando la población se fía. Adicionalmente, la independiente, tratará temas fundamentales para entender a la inteligencia artificial.

Seguidamente, el segundo capítulo presentará el estudio virtual a los estudiantes de 11vo grado del colegio San Agustín de Lima, Perú, quienes han calificado textos generados por la inteligencia artificial. Asimismo, se analizarán los datos obtenidos.

Por último, el trabajo evaluará las ventajas y desventajas de GPT-2. Al igual que, la importancia social y ética, ahondando en puntos como: personas y máquinas, la brecha digital y la igualdad de acceso, y ciudadanía digital. Planteando, finalmente, soluciones.

CAPÍTULO 1 - MARCO TEÓRICO

1.1 La información

1.1.1 Definición del término “información”

¿Existe algún término con gran peso, importancia y repercusión en el flagrante mundo conectado? La respuesta es sí, es “información”. Con el tiempo, estas once letras se han adaptado a cambios. Diferentes definiciones se les han dado según el contexto. El siguiente párrafo aterrizará la definición que será utilizada como referencia para comprender los posteriores subtemas.

De acuerdo con la Real Academia Española (s.f.), información posee ocho significados distintos, siendo apropiada la definición 5: “Comunicación o adquisición de conocimientos que permiten ampliar o precisar los que se poseen sobre una materia determinada”. Por otro lado, la Unesco (2005) menciona que la información es un instrumento de conocimiento, originado por el deseo de intercambiar y hacer más eficaz su transmisión (p. 19). Ser el transmisor del conocimiento es como si todo el progreso recayera sobre una palabra, tanto para bien y mal. En caso algo sea engañoso y el mundo generara conocimiento a partir de ello, desencadenaría consecuencias de gran magnitud (el impacto se retomará a posteriori). No se debe subestimar el poder de una palabra.

1.2 La información ofrecida por la inteligencia artificial GPT-2

1.2.1 Ejemplo y análisis de textos desarrollados por GPT-2

GPT-2 puede generar una variedad de textos a partir de una entrada. La inteligencia artificial, siguiendo la única tarea de predecir la próxima palabra, puede respetar la estructura del texto. Una de las pruebas realizadas por OpenAI, ubicada en Anexos - Ejemplo 1 GPT-2, respalda ello.

En la experimentación, se realizaron seis repeticiones, estudiando la variación en las salidas. En todos los casos, GPT-2 retornó instrucciones, tanto a modo de lista o párrafo. Al leer el texto, es evidente que los investigadores, intencionalmente nombraron diferentes dulces. Y, tras mencionarlos, viraron radicalmente a solo

galletas de merengue. GPT-2, acertadamente, continuó los 6 textos con estas. Además, prosiguió con la última idea, la preparación del dulce.

De manera personal, entre todas las salidas, la más impresionante fue la #4, Anexos - Ejemplo 1 GPT-2. Me pareció extraordinario como continuó la noción final y completó el texto con los ingredientes para preparar las galletas. Inclusive, consideró conceptos mencionados en la entrada, como “Peppermint Jo Jos”. GPT-2 demostró entender el contenido y forma del texto.

1.2.2 La veracidad de los textos ofrecidos por GPT-2

La veracidad de los textos brindados por GPT-2 no es una respuesta binaria, es más extensa. Por ejemplo, la IA puede redactar textos literarios, donde no importa si el contenido es veraz. Además, de acuerdo con OpenAI, empresa desarrolladora, manifiesta que GPT-2 “permite al usuario generar continuaciones realistas y coherentes sobre un tema de su elección” (2019, párr. 5). GPT-2 intenta hacer creer al usuario que la información es real; por lo tanto, no es obligatoriamente fiable, simplemente redacta de manera lógica.

Sin embargo, en algunas ocasiones, la información puede ser real. Esto no sucede adrede, sino porque la IA selecciona las palabras de acuerdo a su peso dentro del modelo. Una verdad objetiva al ser tan nombrada en Internet aumenta la probabilidad de ser mencionada cuando la entrada trata sobre esta. Aun así, la veracidad no está garantizada; ya que, de una misma entrada puede haber diferentes salidas. En definitiva, GPT-2 no devuelve información verídica, su único fin es ser coherente; no obstante, al elaborar textos lógicos, puede engañar al lector fácilmente.

1.3 La información falsa

1.3.1 El impacto de la información falsa en la sociedad

Fundamentalmente, la información influye en la sociedad debido a Internet. La población se informa primordialmente por ahí. Domo en el 2020 expone que cada minuto se enviaban 41 666 667 mensajes por WhatsApp, se subían 147 000 fotos a Facebook, 347 222 publicaciones en Instagram, 479 452 personas estaban en Reddit y 319 usuarios se registraban en Twitter (Anexos - Figura 1). Al haber tantos nodos,

la divulgación es sencilla y masiva. Además, la fuente señala que, en 6 años, la cantidad de usuarios en Internet creció 50%. ¿Qué nos depara en otros 6? ¡¿30?!

No obstante, la información impacta en el saber de la sociedad. Las Naciones Unidas (2005) alude que: “La libre circulación de la información y la disponibilidad de un gran cúmulo de conocimientos pueden facilitar su utilización malintencionada” (p. 125). Imaginar 1% de las cifras enunciadas anteriormente como información falsa es un escenario pavoroso. Que aún no haya acontecido, no significa que no pueda.

Inclusive, la información falsa a manos de la IA puede repercutir en cualquier área. Es posible crear textos políticos, educativos, informativos y más. Hao (2019) resalta que a GPT-2 se le da bien crear noticias falsas (párr. 2). Por lo que, debemos estar alertas; ya vivenciamos al COVID-19, que empezó en China y culminó en pandemia, prevengamos que el siguiente virus sea el de la información falsa.

1.3.2 Antecedentes de difusión de información falsa

El gran canal de distribución que establece Internet sin ninguna discriminación hace recaer toda restricción sobre la ética. Lamentablemente, un incidente donde se vio comprometida fue el de Cambridge Analytica. De acuerdo con Amer y Noujaim (2019), la empresa especializada en datos y algoritmos (como se citó a Nix,12m0s) fue contratada para la campaña de Donald Trump en 2016 (13m30s). Posteriormente, fue investigada junto a Facebook por atentar contra la privacidad de los usuarios (51m40s). Durante las elecciones realizaron encuestas que creaban modelos de personalidad de cada elector, clasificaron a las personas y se enfocaron en las “persuasibles”, incitándolas a votar por los republicanos mediante Facebook (como se citó a Kaiser, 41m15s). Como es de esperar, en muchas ocasiones, la información que divulgaban era falsa. BBC (2018) entrevistó a Wylie (2018), exempleado, quien ante la pregunta si plantaban noticias falsas, respondió afirmando (0m1s). Finalmente, el futuro de la mayor potencia lo definió la información falsa, ganando Trump las elecciones.

1.4 Fundamentos de la Inteligencia Artificial

1.4.1 Definición del concepto “inteligencia artificial”

Inteligencia artificial (IA) es un concepto que en la última centuria ha dado un gran salto en cuanto a popularidad. El surgimiento de la Big Data puso a disposición de científicos e ingenieros un vasto conjunto de datos para el entrenamiento de modelos computacionales.

En los 90 se dio el BOOM, durante la conferencia de Dartmouth (Hardy, 2001, p. 4). Pero la idea nació desde el siglo XVII. Buchanan (2006) alude que el filósofo Wilhelm ya había planteado el concepto de dispositivos mecánicos capaces de razonar mediante la lógica (p. 53).

Por otro lado, Russell y Norvig (2010) definen a la IA en cuatro categorías: “Pensando Humanamente, Actuando Humanamente, Pensado Racionalmente y Actuando Racionalmente” (p. 2). Los apartados no son excluyentes, se complementan. Armonizándolos, se llega a la siguiente definición: la IA es el esfuerzo de hacer a las computadoras pensar a partir del aprendizaje. Siendo el fin que desarrollen actividades requeridas de inteligencia, tales como tomar decisiones o resolver problemas. Además, es el estudio y diseño de modelos computacionales, los cuales permiten a los ordenadores percibir, razonar y actuar, convirtiéndola en un arte. Concretando, la IA es la manera como las computadoras replican el razonamiento humano, una tarea tan compleja y pulcra, digna de un artista.

1.4.2 El aprendizaje automático (ML)

Dentro de la IA, Díaz (2019) resalta el aprendizaje simbólico y automático, machine learning (ML). Del último, surge el Deep Learning (DP), basado en redes neuronales (NN), aquí pertenece GPT-2. Para comprender íntegramente su funcionamiento, primero se abordará el ML. Catalán (2019) declara cuatro pasos: preparación de datos, creación, entrenamiento y testeo del modelo. Primero, el dataset debe acondicionarse a la entrada (es recomendable fraccionarlo, 80% y 20%; evitando sobrecargar los pesos al entrenar). Después, se crea el modelo; escogiendo un algoritmo, función de error y optimizador. Luego, se entrena con el dataset mayor. Finalmente, se testea, fijando de entrada al otro.

Adicionalmente, de acuerdo con Celis (2019), hay diversos aprendizajes:

- Supervisado: consiste en entrenar una IA con los datos de entrada y salida esperada (etiquetas). Se utiliza en clasificación y regresión, siendo la regresión lineal el algoritmo insignia.

- No supervisado: reconoce patrones en un dataset y no precisa entrenamiento etiquetado. Es empleado en agrupamiento, binario o multiclase; aunque puede designarse en reducir la dimensionalidad de datos. Algunos algoritmos son: support vectors machine, k-nearest neighbors y k-means.

- Por refuerzo: se premia o castiga las decisiones del algoritmo. Comúnmente es destinado para interactuar con un entorno. Posee aplicaciones en sistemas de navegación o software dedicados a ganar videojuegos, destacando el Q - learning.

1.4.3 El aprendizaje profundo (DL)

El DL comparte los cuatro pasos del ML, diferenciándose por usar NN. Shiffman (2012) declara: “El cerebro humano puede describirse como una red neuronal biológica: una red interconectada de neuronas que transmiten patrones elaborados de señales eléctricas” (p.444). De esto, se puede inferir que las NN son neuronas artificiales interconectadas que intentan imitar el funcionamiento del cerebro utilizando operaciones matemáticas complejas. Por ello, es idóneo presentar a la más básica, el perceptrón:

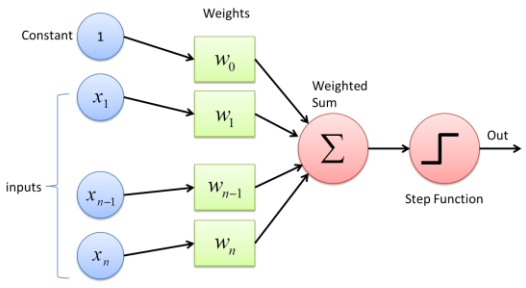


Figura 2 - Perceptrón[1]

[1] Nota. Adaptado de Perceptron [imagen], por S. Sharma, 2017, towards data science (https://towardsdatascience.com/what-the-hell-is-perceptron-626217814f53).

Parafraseando a Amini (2021, 12m28s); primero, la neurona recibe los datos y son multiplicados con sus pesos (pesos sinápticos), atribuyéndole más importancia a ciertos valores. Después, se suman. Luego, el total actúa de entrada para la función de activación, usualmente sigmoide o lineal. Esta establece una "regla" entre los datos; además, es aquí donde otra neurona podría conectarse. Finalmente, retorna un resultado.

En el caso de GPT-2, utiliza una NN conocida como red neuronal convolucional (CNN), caracterizada por usar la operación matemática convolución al filtrar los datos de entrada. LeCun et al. (2015) afirman que estas poseen cuatro claves principales: "conexiones locales, pesos compartidos, agrupación y muchas capas" (p. 439). La basta cantidad de capas supone distintos niveles de NN interconectadas, denominándose inherentemente profunda y permitiendo efectuar arduos cálculos con certeza.

1.4.4 Procesamiento de Lenguaje Natural (NLP)

El Procesamiento de Lenguaje Natural (NLP, por sus siglas en inglés) es una subrama del ML. En este se basa GPT-2. De acuerdo con Gelbukh (2010), por NLP "se entiende la habilidad de la máquina para procesar la información comunicada" (p. 6). Así como nos comunicamos con otras personas mediante nuestro idioma, ahora podemos listar a las computadoras.

Tal como, un bebé no sabe gramática cuando aprende a hablar, sino que analiza las palabras y las relaciona con lo conocido (Gelbukh, 2010, p. 6); lo mismo hacen las computadoras. Santana (2019) como citó a Ruder (2018), expresa que dividen oraciones en caracteres, palabras o sub-palabras (9m5s); esta última utiliza GPT-2. Después, se usa One Hot Encoding, creando para cada token una matriz del tamaño del vocabulario y marcando numéricamente su posición (9m30s). Sin embargo, teóricamente, no todas las palabras tienen mismo vínculo. Por ello, se efectúan word embeddings, asignándoles magnitudes mediante ML (10m55s). Así, la computadora comprende su significado y gramática, permitiendo crear un modelo de lenguaje como GPT-2 (14m5s).

Por otro lado, en Perú, grandes empresas ya están usando esta tecnología. Un ejemplo es el asistente virtual de Movistar, el cual atiende "todas las llamadas de sus

clientes” (Agencia EFE, 2019, párr. 2). Este reconoce y procesa la voz de las personas siguiendo la explicación superior.

1.5 La inteligencia artificial GPT-2

1.5.1 Información general

El segundo modelo Transformador Generador Pre entrenado o GPT-2, es una IA creada por OpenAI y basada en NLP. Tiene el objetivo de predecir la siguiente palabra a partir de cierta entrada. Además, conforme con OpenAI (2019), la versión más grande del modelo posee 1,5 millones de parámetros (párr. 2), igual a 8 millones de páginas web o 40 GB de texto (párr. 1 - 2). Todo extraído con web scrapers en Reddit, filtrando los sitios por puntuación de +3 karmas; debido a ello, el conjunto de datos posee variedad y compostura (Radford et al., 2019, p. 3). Además, al ser pre entrenado puede transferir conocimiento; por lo que, GPT-2 podría ser la base de otro modelo, simplificando su entrenamiento (Santana, 2019, 15m18s).

Un ejemplo de GPT-2 como modelo pre entrenado es el Proyecto Trevor. Este busca ayudar a los jóvenes LGBTQ con pensamientos suicidas durante su proceso de aceptación. Consta de un chatbot llamado Riley y de acuerdo con Ohlheiser (2021), “el chatbot usa GPT-2 para sus habilidades básicas de conversación” (párr. 17). Gracias a su conocimiento del idioma inglés, simplifica la ejecución de propuestas transcendentes. Lo que empezó como un modelo que predice la siguiente palabra, ahora salva vidas.

1.5.2 Funciones extras de GPT-2

Sorprendentemente, GPT-2 solo prediciendo la siguiente palabra desarrolló otras aptitudes, volviéndose un modelo multitarea. A continuación, las funciones extras según Radford et al. (2019):

1. Respuesta a preguntas: le brindaron “pares de preguntas y respuestas de ejemplo”, dejando al final una sin respuesta. GPT-2 fue inferior a modelos especializados. Pero ante: “¿Quién escribió el libro El origen de las especies?”, continuó “Darwin” un 83,4% de las veces (p. 7).

2. Traducción: se utilizaron pares de ejemplo: [oración en inglés] = [oración en otro idioma]. GPT-2 obtuvo 11.5 BLEU, resultado alentador, pero menor a otros modelos. Sin embargo, tradujo francés con tan solo 10 MB de datos, mientras aquellos en inglés eran 500x mayor (pp. 6 - 7).

3. Compresión lectora: se colocó un texto seguido de pares de preguntas y respuestas. Lo obtenido fue “de vanguardia”. En cuentos infantiles, comprendió 93.3% de los sustantivos comunes y 89,1% de las entidades nombradas; mientras el ser humano 95% y 92,5%, respectivamente (pp. 5 - 6).

4. Generación de resúmenes: en Internet, cuando un texto es extenso, se escribe TL; DR, haciendo hincapié al resumen. Si se colocaba al final de una entrada, GPT-2 lo entendía. Y, usándolo obtuvo un puntaje promedio de 21.4 ROUGE F1, pero caía en 6 cuando no (p. 6).

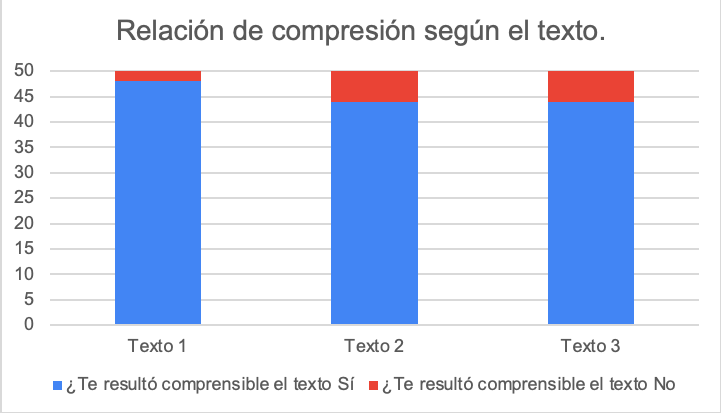
CAPÍTULO 2 - ANÁLISIS ESTADÍSTICO

2.1 Evaluación de los textos generados por GPT-2

Se elaboró una encuesta en Google Forms a 50 alumnos, 25 hombres y 25 mujeres, de 11vo grado del colegio San Agustín de Lima, Perú con el objetivo de calificar los textos continuados por GPT-2. El formulario (Anexos - Encuesta) posee 3 textos traducidos del informe de Radford et al., cada uno con 5 preguntas.

2.1.1 Gráfica 1 - Relación de compresión según el texto

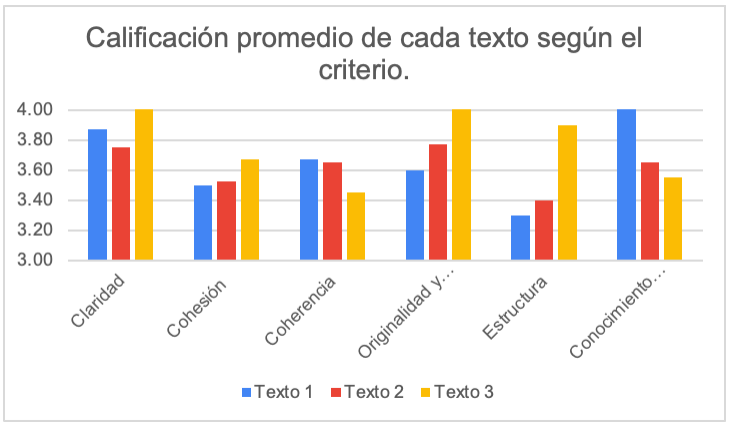
En relación con la primera pregunta, en promedio, 9/10 personas comprenden los textos. Destacó el #1, 95% de los encuestados respondió positivamente. En los restantes, 87,5% lo comprendieron. Es probable que el primero sobresalió por tener párrafos cortos, enfocando eficazmente la idea principal.



Gráfica 1

2.1.2 Gráfica 2 - Calificación promedio de cada texto según el criterio

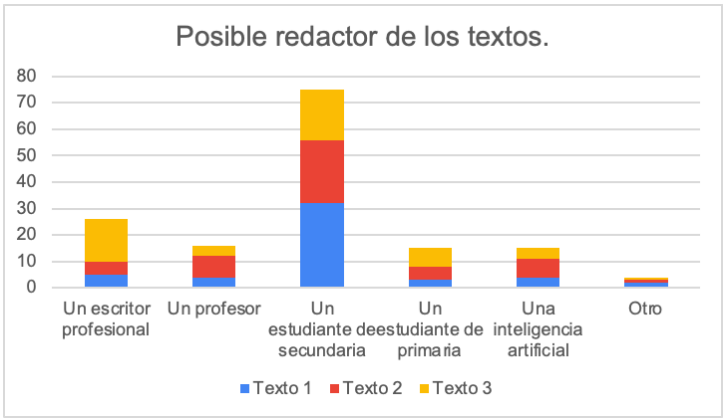
Para la segunda, el texto mejor calificado es el tercero, obtuvo 22,65/30 puntos en promedio. Antagónicamente, está el segundo con 21,75, siendo la media 22,13. El criterio renombrado es: claridad, aunque en todos se consiguió +3,30/5. Los resultados demuestran que GPT-2 es competente.



Gráfica 2

2.1.3 Gráfica 3 - Posible redactor de los textos

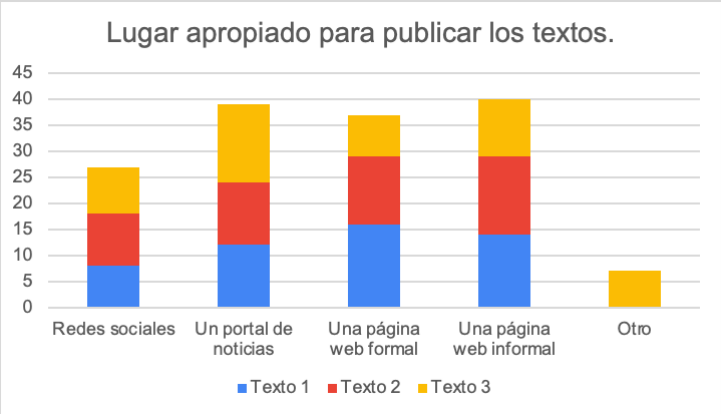
En cuanto a la tercera gráfica, la mayoría afirma que los textos fueron escritos por un estudiante de secundaria o un escritor profesional. De 10 personas, aproximadamente 8 marcaron la primera, 2,88 veces la otra, y menos de 2 la opción “IA”. Las respuestas evidencian el potencial de GPT-2 aparentando ser humano.



Gráfica 3

2.1.4 Gráfica 4 - Lugar apropiado para publicar los textos

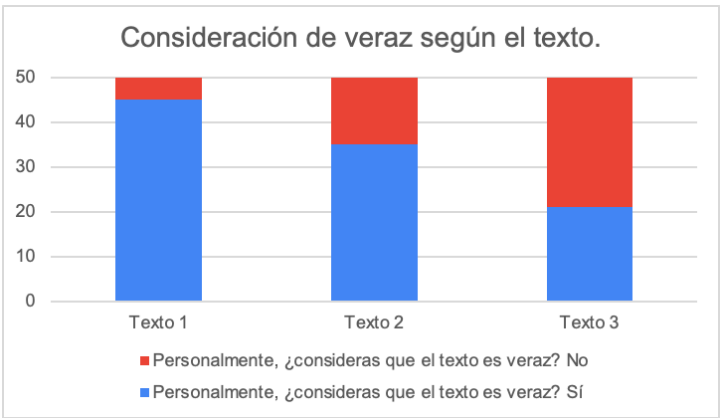
Sobre la cuarta, se concluye que, principalmente, los textos calzan en un portal de noticias o página web informal. Asimismo, las personas dudan ante personajes mágicos, aludiendo que podría ser un cuento, blog, video informal o redacciones de Wattpad. GPT-2 demostró versatilidad, escribiendo tanto textos informativos como informales.



Gráfica 4

2.1.5 Gráfica 5 - Consideración de veraz según el texto

Analizando la última interrogante, la diferencia entre quienes consideran veraz un texto no surrealista (Texto 1) y otro que sí (Texto 3) es de 48%. Los resultados demuestran que la mayoría identifica textos falsos fantasiosos, aunque preocupa la fiabilidad en los otros.



Gráfica 5

En definitiva, los textos continuados por GPT-2 son considerados veraces. Los entrevistados entendieron y calificaron los textos superior a la media de la banda de calificación. También, la mayoría consideró que fueron escritos por personas, menos de 2/10 identificaron que intervino una IA. Por último, al considerar que fueron escritos por humanos adoptaron una postura de confianza, señalando que podrían publicarse en plataformas cotidianas y apoyando fuertemente su fiabilidad cuando no presentaban personajes mágicos.

CAPÍTULO 3 - CUESTIONES SOCIALES Y ÉTICAS

3.1 Ventajas y desventajas

Ventajas

1. Las personas serían más eficientes trabajando cooperativamente con la IA. Jiménez (2017) citando a Harmsen (2017) alude que “la cooperación entre humanos y máquinas permitirá que ambas partes mejoren” (párr. 9). GPT-2 posee un desempeño de +3,30/5, es factible una simbiosis con escritores. Ellos elaborarían la introducción, GPT-2 continua y finalmente lo corrigen.

2. La IA se adapta. Según el Parlamento Europeo (2021): “Los sistemas de IA son capaces de adaptar su comportamiento en cierta medida, analizar los efectos de acciones previas y de trabajar de manera autónoma” (párr. 4). GPT-2 es un modelo pre entrenado y puede adecuarse a cualquier tarea de texto y estructura o a redacciones específicas.

Desventajas

1. Mejorar una IA requiere gran cantidad de datos. Lee (2019) comenta que el avance de esta tecnología se basa en el consumo de datos (párr. 56). Medrar GPT-2 y sus predecesores precisará de miles de millones de data points extras, demandando adicionalmente un mejor hardware. La inversión es alta y los recursos finitos, el desarrollo de IA es limitado.

2. GPT-2 facilita la creación de información falsa. Merino (2020) expone: “sus desarrolladores habían anunciado que no publicarían la versión completa de la misma, por miedo a un ‘mal uso’” (párr. 1). Inicialmente, publicaron una versión inferior a 1,5 millones de parámetros, temían oleadas de contenido engañoso. La IA es peligrosa hasta para sus creadores.

3.2 Importancia social y ética

Personas y máquinas

La perfección de algoritmos permite la automatización con IA. Un ejemplo es la conducción; antes era una persona, ahora una computadora. Rus (2020) comenta

que la empresa de taxis Waymo, “entre el 5% y el 10% de los viajes que han hecho en 2020 han sido completamente sin conductor” (párr. 2). La conducción autónoma supera en periodo de desarrollo al NLP, solo es cuestión de tiempo para que un nuevo GPT-2 iguale al humano.

La brecha digital y la igualdad de acceso

El conocimiento de avances en IA está restringido al acceso a Internet. La “Inteligencia artificial (IA) y cloud computing son dos tecnologías íntimamente relacionadas” (Agencia B12, 2020, párr. 1). Innovaciones de código abierto como GPT-2 son publicadas usualmente en páginas webs. Asimismo, software como los asistentes virtuales precisan de la red y no todos poseen conexión.

Ciudadanía digital

GPT-2 supone un aumento en la adopción de información falsa. Diazgranados (2020) citando al estudio de Kaspersky (2020) menciona que: “70% de los latinoamericanos no sabe detectar o no está seguro de reconocer en Internet una noticia falsa de una verdadera” (párr. 3). Actualmente, la población puede ser engañada sin problemas. Latinoamérica no está preparada para encarar a los modelos de lenguaje, la educación en reconocimiento de información engañosa urge.

3.3 Soluciones

1. El estado debería brindar el acceso a Internet a todos. La asamblea general de la ONU para el año 2016 declaró como derecho humano el acceso a Internet (Barry, 2020, párr. 3). Sin embargo, los países no están sujetos a sanciones. La brecha digital en el conocimiento de tecnologías como GPT-2 se acortaría si se impusiera con firmeza lo expresado por la ONU hace 5 años.

2. Motivar el trabajo en la tecnología y salir del confort. Así como en la segunda revolución industrial los agricultores migraron a las fábricas (Goos, 2013, p. 1), similarmente sucederá con los trabajadores actuales. Por ejemplo, los escritores, en un futuro, serán reemplazados por IA como GPT-2. Pero serán indispensables en el perfeccionamiento de los textos, laborando con los programadores. El cambio no es negativo, el no adaptarse lo es.

3. Crear campañas sobre identificación de contenido falso. Una a imitar es: Juntos contra la desinformación sobre COVID-19, por la OMS y Facebook. Esta afirma que “las noticias falsas pueden manipular los sentimientos con un clic” (Agencia Europa Press, 2021, párr. 5). En el presente, la información falsa ya supone un problema. No debemos esperar ser afectados por la viralización de información engañosa redactada con IA, debemos prevenirlo formando una ciudadanía latinoamericana digital.

CONCLUSIÓN

En definitiva, GPT-2 mediante el NLP influye en gran medida en la valoración de sus textos como veraces, por parte de los estudiantes de 11vo grado del colegio San Agustín de Lima, Perú. La IA demostró ser comprensible, competente, versátil y capaz de engañar a las personas. Sus textos fueron considerados fiables, aunque con tendencia negativa cuando fantaseaban.

Paralelamente, el impacto de la información falsa es colosal, perjudicando el conocimiento de la sociedad y logrando hasta victorias en elecciones presidenciales.

Lamentablemente, con GPT-2, la producción de información falsa es sencilla. La IA omite la veracidad, simplemente busca coherencia, prediciendo la siguiente palabra según las anteriores. Además, el NLP es una tecnología nueva a comparación de otras ramas del ML. Si en la actualidad ya posee un gran desempeño, el futuro es amenazador.

Por último, GPT-2 adeuda importancia social y ética. Al operar mediante Internet, acarrea una brecha digital en el acceso. Debido a ello, el estado debería brindarlo a todos. También, GPT-2 supone reemplazar puestos laborales; sin embargo, las personas no serán obsoletas. Finalmente, la ciudadanía latinoamericana debe prepararse para encarar a GPT-2, ejecutar campañas contra la información falsa es fundamental.

Concretando, el NLP crece rápidamente y debemos ser veloces. Pero esto recién comienza y tenemos que estar atentos ante un ascenso inminente. Aún estamos a tiempo de prevenir una pandemia de información falsa producida por futuros GPT-2.

BIBLIOGRAFÍA

Agencia B12. (2020). *Cómo la Inteligencia Artificial está mejorando el Cloud Computing en las empresas.* B12 Tech4Business. Recuperado de https://agenciab12.com/noticia/como-inteligencia-artificial-esta-mejorando-cloud-computing-empresas

Agencia EFE. (2019). *Movistar atenderá llamadas de clientes de Perú con inteligencia artificial.* Gestión. Recuperado de https://gestion.pe/tecnologia/movistar-atendera-llamadas-clientes-peru-inteligencia-artificial-267875-noticia/

Agencia Europa Press. (2021). *La nueva campaña de Facebook y la OMS enseña a detectar noticias falsas sobre el COVID-19.* El Comercio. Recuperado de https://elcomercio.pe/tecnologia/actualidad/la-nueva-campana-de-facebook-y-la-oms-ensena-a-detectar-noticias-falsas-sobre-el-covid-19-noticia/

Amer, K. y Noujaim, J. (2019). *Nada es privado* [documental, video en línea]. Netflix Originals. Recuperado de https://www.netflix.com/watch/80117542

Amini, A. (2020). *MIT Introduction to Deep Learning 6.S191* [archivo de video]. YouTube. Recuperado de https://www.youtube.com/watch?v=5tvmMX8r_OM

Barry, J. (2020). *La COVID-19 demuestra por qué el acceso a Internet es un derecho humano.* OpenGlobalRights. Recuperado de https://www.openglobalrights.org/covid-19-exposes-why-access-to-internet-is-human-right/?lang=Spanish

BBC. (2018). *Cambridge Analytica planted fake news* [Cambridge Analytica plantó noticias falsas] [archivo de video]. YouTube. Recuperado de https://www.youtube.com/watch?v=mjtR3W3eAFU

Buchanan, B. G. (2005). *A (Very) Brief History of Artificial Intelligence* [Una (breve) historia de la inteligencia artificial]. AI Magazine, 26(4), 53. Recuperado de https://doi.org/10.1609/aimag.v26i4.1848

Catalán, A., Celis, R. y Torres, D. (2019). *Curso de Introducción a Machine Learning*. Platzi. Recuperado de https://platzi.com/clases/machine-learning-2019/

Díaz, R. (2019). *Curso de Fundamentos Matemáticos para Inteligencia Artificial*. Platzi. Recuperado de https://platzi.com/clases/matematicas-ai-2019/

Diazgranados, H. (2020). *70% de los latinoamericanos desconoce cómo detectar una fake news. Kapersky daily*. Recuperado de https://latam.kaspersky.com/blog/70-de-los-latinoamericanos-desconoce-como-detectar-una-fake-news/17015/

Domo (2020). *Data Never Sleeps 8.0* [Los datos nunca duermen 8.0] [infografía, archivo de imagen]. Domo. Recuperado de https://www.domo.com/learn/infographic/data-never-sleeps-8

Gelbukh, A. (2010). *Procesamiento de Lenguaje Natural y sus Aplicaciones*. Komputer Sapiens, pp. 6 - 11, vol. 1. Recuperado de https://www.gelbukh.com/CV/Publications/2010/Procesamiento%20de%20lenguaje%20natural%20y%20sus%20aplicaciones.pdf

Goos, M. (2013). *Cómo está cambiando el mundo del trabajo:análisis de los datos* [versión PDF]. Organización Internacional del Trabajo, ed. 1. Recuperado de https://www.ilo.org/public/spanish/dialogue/actemp/downloads/events/2013/symp/how_worldofwork_changing_sp.pdf

Hao, K. (2019). *Historia de la IA que domina las 'fake news' y su impacto en la sociedad.* MIT Techonology Review. Recuperado de https://www.technologyreview.es/s/11412/historia-de-la-ia-que-domina-las-fake-news-y-su-impacto-en-la-sociedad

Hardy, T. (2001). *IA: Inteligencia artificial.* Polis, Revista de la Universidad Bolivariana, pp. 0 - 23, vol. 1 Recuperado de https://www.redalyc.org/pdf/305/30500219.pdf

Jiménez, M. (2017). *Harmsen (Google): "La cooperación entre humanos y máquinas permitirá que ambos mejoren".* El País. Recuperado de https://elpais.com/retina/2017/04/09/tendencias/1491755541_042969.html

LeCun, Y., Bengio, Y. e Hinton, G. (2015). *Deep Learning* [Aprendizaje Profundo] [versión PDF]. Universidad de Toronto, pp. 436 - 444, vol. 521. Recuperado de https://www.cs.toronto.edu/~hinton/absps/NatureDeepReview.pdf

Lee, K. (2019). *La inteligencia artificial y el futuro del trabajo: una perspectiva china.* OpenMind BBVA. Recuperado de https://www.bbvaopenmind.com/articulos/inteligencia-artificial-y-futuro-del-trabajo-perspectiva-china/

Merino, M. (2019). *GPT-2: qué sabemos y qué no del generador de textos con IA que OpenAI dice haber censurado por ser demasiado peligroso.* Xataka. Recuperado de https://www.xataka.com/inteligencia-artificial/gpt-2-que-sabemos-que-no-generador-textos-ia-que-openai-dice-haber-censurado-ser-demasiado-peligroso

Ohlheiser, A. (2021). *Una IA basada en GPT-2 ofrece ayuda a adolescentes LGTBQ en crisis.* MIT Technology Review. Recuperado de

https://www.technologyreview.es/s/13214/una-ia-basada-en-gpt-2-ofrece-ayuda-adolescentes-lgtbq-en-crisis

OpenAI. (2019). *Better Language Models and Their Implications* [Mejores modelos del lenguaje y sus implicaciones]. OpenAI. Recuperado de https://openai.com/blog/better-language-models/

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. Unesco. (2005). *Hacia las sociedades del conocimiento.* Unesdoc, pp. 1 - 244. Recuperado de https://unesdoc.unesco.org/ark:/48223/pf0000141908

Parlamento Europeo. (2020). *¿Qué es la inteligencia artificial y cómo se usa?.* Noticias Parlamento Europeo. Recuperado de https://www.europarl.europa.eu/news/es/headlines/society/20200827STO85804/que-es-la-inteligencia-artificial-y-como-se-usa

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D y Sutskever, I. (2019). *Language Models are Unsupervised Multitask Learners* [Los modelos del lenguaje son aprendices multitarea no supervisados]. OpenAI. Recuperado de https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

Real Academia Española. (s.f.). *Información*. Diccionario de la lengua española. Recuperado de https://dle.rae.es/información

Rus, C. (2020). *El servicio de taxis autónomos de Waymo ya es público: ofrecerán viajes con coches que se conducen totalmente solos.* Xataka. Recuperado de https://www.xataka.com/vehiculos/servicio-taxis-autonomos-waymo-publico-ofreceran-viajes-coches-que-se-conducen-totalmente-solos

Ruseell, S. y Norving, P. (2010). *Artificial Intelligence A Modern Approach* [Inteligencia artificial: un enfoque moderno] [libro, versión pdf]. Pearson, 3. Recuperado de https://zoo.cs.yale.edu/classes/cs470/materials/aima2010.pdf

Santana, C. (2019). *GPT-2 El Impresionante Generador de Texto Censurado* [archivo de video]. YouTube. Recuperado de https://youtu.be/JVT7ITIZY7I

Shiffman, D. (2012). *The Nature of Code* [La naturaleza del código] [libro, versión PDF]. ISBN, ed. 5. Recuperado de http://wtf.tw/ref/shiffman.pdf

ANEXOS

Ejemplo 1 GPT-2

(Radford et al., 2019, p. 19):

Entrada (traducción de inglés a español):

A Aaron le encanta el pastel de chocolate con menta, pero requiere que se combine con mini chispas de chocolate, así que arrojé algunas de ellas entre las capas. También tenía algunas Peppermint Jo Jos a mano, así que las aplasté y arrojé algunas de ellas junto con algunas galletas de merengue trituradas porque, ¿por qué no? Es una mezcla heterogénea de bondad crujiente de galletas con chispas de chocolate y menta. No medí la cantidad de cada cobertura que usé, pero después de probar el producto terminado, desearía haber agregado más. Puede agregar lo que quiera: bastones de caramelo triturados, corteza de menta, etc. Y no tenga miedo de usar mano dura. Textura = buena. Si no tiene moldes para pasteles de 7 pulgadas, puede obtener 3 capas más cortas de 8 pulgadas de este

Salida #4 (traducción de inglés a español):

receta.

Galletas de merengue

1 3/4 taza de mantequilla ablandada

1 taza de azúcar

1 yema de huevo

3 cucharadas de crema espesa

3 1 / 2-4 tazas de harina

pizca de sal

peppermint Jo Jos

1 1/4 tazas de azúcar en polvo

1 taza de nueces picadas

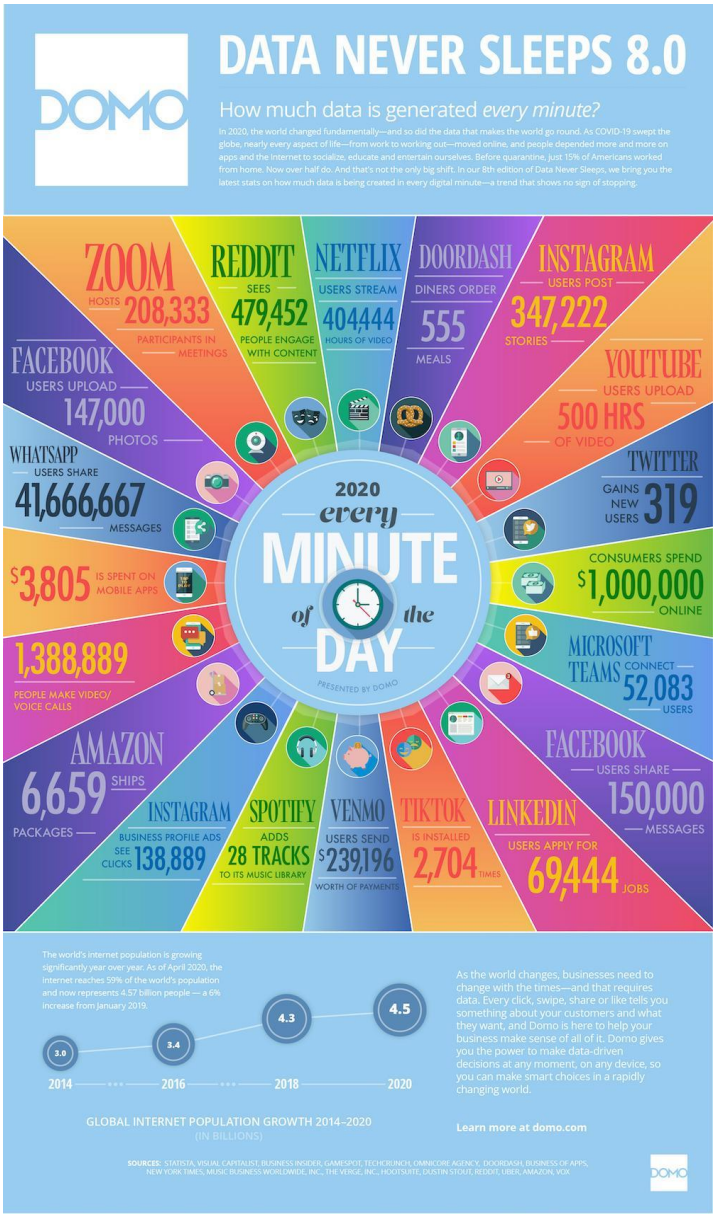
1/2 taza de hojas de menta finamente picadas

1/2 taza de menta fresca picada (aproximadamente 1/2 hoja)

1 cucharadita extracto de menta

1/4 cucharadita pizca de sal marina de mar

Figura 1 - Infografía “Data Never Sleeps 8.0”



Nota. Adaptado de Data Never Sleeps 8.0 [infografía], por Domo, 2020, Domo (https://www.domo.com/learn/infographic/data-never-sleeps-8).

Encuesta para los estudiantes de 11vo grado del colegio San Agustín de Lima, Perú

Enlace al formulario → https://forms.gle/dAjAEdv2FCDq7tet6

Lea con atención los siguientes textos traducidos y responda las preguntas adyacentes.

Fuente de los textos:

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D y Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners [Los modelos del lenguaje son aprendices multitarea no supervisados]. OpenAI. Recuperado de https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

Texto 1 - Artículo (Radford et al., 2019, p. 21)

El hombre prehistórico dibujó una increíble variedad de bestias prehistóricas en las toscas paredes de piedra caliza de una cueva en la Francia actual hace 36.000 años.

Ahora, con la ayuda de tecnología de punta, esas obras de arte de la cueva Chauvet-Pont-d'Arc se han reproducido para crear la réplica de la cueva más grande del mundo.

La caverna artificial llamada Caverne du Pont-d'Arc se ha construido a pocas millas del sitio original en Vallon-Pont-D'arc en el sur de Francia y contiene 1,000 dibujos cuidadosamente reproducidos, así como alrededor de 450 huesos y otras características…

Hombres y mujeres de las cavernas esbozaron una increíble variedad de bestias prehistóricas en las ásperas paredes de piedra caliza de una cueva hace 36.000 años y ahora se ha creado una réplica (en la foto)

El sitio original en Vallon-Pont-D'arc en el sur de Francia es un sitio del Patrimonio Mundial de la Unesco y es la cueva más antigua conocida y mejor conservada decorada por el hombre. La réplica de la cueva se construyó a pocos kilómetros del sitio original en Vallon-Pont-D'Arc en el sur de Francia. La cueva contiene imágenes de 14 especies diferentes de animales, incluidos rinocerontes lanudos, mamuts y grandes felinos.

Texto 2 - Artículo (Radford et al., 2019, p. 21)

"Si puedes encontrar mi submarino, es tuyo", dijo una vez el multimillonario petrolero ruso Roman Abramovich.

Y parece que el oligarca y propietario del Chelsea FC, cuyo submarino es solo uno de los extras que vino con su superyate Eclipse de £ 300 millones (perfecto para llegar a la costa sin ser detectado), no es el único hombre de negocios rico que se dedica a la exploración submarina.

Apodados como la verdadera "última frontera" de la Tierra, los océanos están todavía tan poco explorados que los multimillonarios hacen cola para comprar embarcaciones que les permitan vislumbrar los espectaculares paisajes marinos y la increíble vida salvaje de los océanos del mundo.

Entonces, si tiene unos pocos millones de sobra en el banco y desea algunas instantáneas navideñas y publicaciones de Instagram que realmente superen a todos los demás, es posible que desee comprar una de estas...

Whale of a time: el OrcaSub lo lleva a 2000 pies debajo de la superficie en dos domos de observación de Perspex presurizados para una exploración óptima.

El submarino U-Boat Worx puede sumergirse a 300 metros (984 pies) y es capaz de sumergirse hasta el fondo del océano. Se puede utilizar para bucear, hacer esnórquel e incluso bucear con esnórquel. El submarino también está equipado con una plataforma de snorkel.

Texto 3 - Noticia (Radford et al., 2019, p. 20)

En un hallazgo impactante, un científico descubrió una manada de unicornios que vivían en un valle remoto, previamente inexplorado, en la Cordillera de los Andes. Aún más sorprendente para los investigadores fue el hecho de que los unicornios hablaban un inglés perfecto.

El científico nombró a la población, por su cuerno distintivo, Unicornio de Ovidio. Estos unicornios de cuatro cuernos, de color blanco plateado, eran previamente desconocidos para la ciencia.

Ahora, después de casi dos siglos, el misterio de lo que provocó este extraño fenómeno está finalmente resuelto.

El Dr. Jorge Pérez, biólogo evolutivo de la Universidad de La Paz, y varios compañeros, estaban explorando la Cordillera de los Andes cuando encontraron un pequeño valle, sin otros animales ni humanos. Pérez notó que el valle tenía lo que parecía ser una fuente natural, rodeado por dos picos de roca y nieve plateada.

Pérez y los demás luego se aventuraron más adentro del valle. “Para cuando llegamos a la cima de un pico, el agua se veía azul, con algunos cristales en la parte superior”, dijo Pérez.

Pérez y sus amigos se sorprendieron al ver la manada de unicornios. Estas criaturas se podían ver desde el aire sin tener que moverse demasiado para verlas, estaban tan cerca que podían tocar sus cuernos.

Al examinar estas extrañas criaturas, los científicos descubrieron que las criaturas también hablaban un inglés bastante regular. Pérez afirmó: “Podemos ver, por ejemplo, que tienen un común "Lenguaje", algo así como un dialecto o dialéctica".

El Dr. Pérez cree que los unicornios pueden haberse originado en Argentina, donde se creía que los animales eran descendientes de una raza perdida de personas que vivían allí antes de la llegada de los humanos a esas partes de América del Sur.

Si bien sus orígenes aún no están claros, algunos creen que quizás las criaturas fueron creadas cuando un humano y un unicornio se conocieron en una época

anterior a la civilización humana. Según Pérez, “en América del Sur, este tipo de incidentes parecen ser bastante comunes”.

Sin embargo, Pérez también señaló que es probable que la única forma de saber con certeza si los unicornios son de hecho descendientes de una raza alienígena perdida sea a través del ADN. “Pero parecen poder comunicarse bastante bien en inglés, lo que creo que es un signo de evolución, o al menos un cambio en la organización social”, dijo el científico.

Preguntas para cada texto:

1. ¿Te resultó comprensible el texto?
    - Sí
    - No
2. Tras leer el texto, cómo calificarías los siguientes criterios:
    - Claridad
    - Cohesión
    - Coherencia
    - Originalidad y Creatividad
    - Estructura
    - Conocimiento sobre el tema

3. Consideras que el texto pudo ser redactado por:

    - Un escritor profesional
    - Un profesor
    - Un estudiante de secundaria
    - Un estudiante de primaria
    - Una inteligencia artificial
    - Otro:

4. Consideras que el texto es apropiado para:

    - Redes sociales
    - Un portal de noticias
    - Una página web formal
    - Una página web informal
    - Otro:

5. Personalmente, ¿consideras que el texto es veraz?:

    - Sí
    - No