



Prototipos de procesamiento de lenguaje natural bajo el modelo de técnicas de aprendizaje automático

Natural language processing prototypes under the model of machine learning techniques

 <https://doi.org/10.47230/Journal.TechInnovation.v2.n2.2023.34-43>

Recibido: 11-08-2023


Aceptado: 11-10-2023

Publicado: 01-12-2023

María José Ordóñez Palma^{1*}

 <https://orcid.org/0009-0008-8568-4275>

Jimmy Leonardo Gutiérrez García²

 <https://orcid.org/0000-0003-2166-5856>

1. Ingeniero en Tecnologías de la Información; Bachiller en Ciencias Únicas Universidad Estatal del Sur de Manabí; Jipijapa, Ecuador.
2. Ingeniero en Computación y Redes; Magíster en Gerencia Educativa; Candidato a Doctor en Tecnologías de la Información y Comunicaciones; Docente Investigador de la Carrera de Tecnologías de la Información de la Facultad de Ciencias Técnicas de la Universidad Estatal del Sur de Manabí; Jipijapa, Ecuador.

Volumen: 2

Número: 2

Año: 2023

Paginación: 34-43

URL: <https://revistas.unesum.edu.ec/JTI/index.php/JTI/article/view/49>

***Correspondencia autor:** ordonez-maria3925@unesum.edu.ec

RESUMEN

El procesamiento del lenguaje natural es una rama de la inteligencia artificial que ha ido ganando protagonismo en el mundo globalizado, debido a que, gracias a dichas técnicas, en la actualidad se pueden realizar traducciones automáticas, revisiones de ortografía o conteo de palabras. Este proyecto plantea estudiar los diferentes prototipos de procesamiento de lenguaje natural con tecnología GPT bajo el modelo de técnicas de aprendizaje automático, plasmando la información en un artículo científico que contribuirá con el avance de este tema, lo cual permitirá a los expertos en el área y desarrolladores, conocer las características de los mismos. Se emplearon los métodos; descriptivo, correlacional, experimental, de revisión bibliográfica e inducción y deducción. Como resultado del estudio, se pudo determinar que, el prototipo BERT representado por la compañía Google, demuestra una AI más eficiente en comparación a otros modelos, teniendo un método de predicción más preciso y auto educándose al momento de aprender las búsquedas que realizan los usuarios, de forma inmediata al devolver una respuesta.

Palabras clave: Procesamiento del lenguaje natural; prototipo; tecnología GPT.

ABSTRACT

Natural language processing is a branch of artificial intelligence that has been gaining prominence in the globalized world, due to the fact that, thanks to these techniques, automatic translations, spell checks or word counts can now be performed. This project proposes to study the different prototypes of natural language processing with GPT technology under the model of machine learning techniques, translating the information into a scientific article that will contribute to the advancement of this topic, which will allow experts in the area and developers, learn about their characteristics. The methods were used; descriptive, correlational, experimental, bibliographic review and induction and deduction. As a result of the study, it was possible to determine that the BERT prototype represented by the Google company demonstrates a more efficient AI compared to other models, having a more accurate prediction method and educating itself when learning the searches made by users, immediately upon returning a response.

Keywords: GPT technolog; natural language processing; prototype.



Creative Commons Attribution 4.0
International (CC BY 4.0)

Introducción

El Procesamiento de Lenguaje Natural (PLN) ha emergido como un campo fascinante y revolucionario que conecta la brecha entre la inteligencia artificial y la comunicación humana. En este contexto, los modelos de lenguaje como GPT (Generative Pre-trained Transformer) han demostrado ser un hito trascendental, impulsando los límites del PLN hacia horizontes antes inimaginables.

Los prototipos de PLN, impulsados por la tecnología GPT y respaldados por las potentes técnicas de aprendizaje automático, Con una capacidad excepcional para aprender patrones, estructuras gramaticales y semánticas, estos prototipos se han convertido en la base de diversas aplicaciones que han transformado nuestra vida cotidiana.

Se examina cómo el aprendizaje automático ha potenciado la adaptabilidad y generalización de los prototipos GPT, permitiéndoles abordar con eficacia la diversidad de idiomas y dialectos que enriquecen la comunicación humana en todo el mundo, entre el Procesamiento de Lenguaje Natural y el aprendizaje automático, destacando el papel fundamental de la tecnología GPT en esta ecuación. Al comprender cómo estos prototipos han transformado nuestra interacción con las máquinas y han hecho que la inteligencia artificial sea más accesible e intuitiva, estaremos mejor preparados para enfrentar los desafíos y oportunidades que el futuro de esta fascinante disciplina nos tiene reservados.

Estas herramientas de apoyo se utilizan para reconocer patrones en los datos y crear reglas basadas en grandes cantidades de información, que no pueden procesarse de forma manual. Algunas problemáticas evidenciadas en la educación brindan un gran potencial investigativo para aplicar la IA, específicamente, desde los procesos de enseñanza-aprendizaje, en los que es posible potencializar su aplicabilidad y desarrollar técnicas automatizadas (Maldonado et al., 2023).

Desarrollo

Un prototipo en sentido lógico es una implementación parcial pero concreta de una parte del sistema, que esencialmente se crea a partir de aspectos muy diversos para explorar el desarrollo del mismo. En relación a la interfaz de usuario, se crean prototipos con el fin de explorar los aspectos interactivos del sistema, incluyendo la accesibilidad, usabilidad y funcionalidad del mismo (Sayago, 2023).

El uso de prototipos en el desarrollo de sistemas no es limitado sólo para probar las interacciones que los usuarios realizan, sino que son muy útiles al momento de realizar otras actividades que se hacen durante el proceso (Sayago, 2023).

Característica

- Las principales características de los prototipos, son las siguientes (Sayago, 2023):
- Son herramientas formidables de la comunicación y participación.
- En la comunicación, se basa en los componentes del equipo y el usuario.
- En la participación, se integra de forma activa en los usuarios del desarrollo.
- Brindan soporte a los diseñadores al momento de elegir entre diversas alternativas.
- Permiten a los diseñadores, explorar diferentes conceptos del diseño antes de determinar los que son definitivos.
- Permiten la evaluación del sistema desde las primeras fases en el desarrollo.
- Son fundamentales para documentar, tanto conceptos funcionales como tareas concretas.
- Son el primer paso para que las ideas sean visibles, concretas y testeables.

- Fomentan la interactividad de las actividades.
- Mejoran la completitud y calidad de las especificaciones funcionales en el sistema.
- Son herramientas con propósito general, sirviendo para comprobación de la fiabilidad técnica de una idea, clarificando los requisitos que quedan como indeterminados.

Tipos de prototipos

- Una representación de lo que va a ser el producto, ya sea en el ordenador o en formato 3D
- La fabricación idéntica del producto final
- Que es el procesamiento de lenguaje natural

Lenguaje natural

El procesamiento del lenguaje natural (PLN) es una tecnología que se emplea para que las máquinas puedan entender el lenguaje de los seres humanos de la misma manera que lo hacen ellos mismos (Inbenta, 2022). El PLN es un subcampo de la inteligencia artificial que combina la informática y lingüística con el fin de interpretar el lenguaje humano, para esto, se descompone el lenguaje en palabras y textos, derivando el contexto de las relaciones entre las palabras y estructura, para extraer el significado (Inbenta, 2022).

Es importante, dado que, como seres humanos se utiliza a diario el lenguaje natural como un medio de comunicación con los demás individuos, lo hacemos gracias a nuestras habilidades que tenemos para procesar, comprender, entender y emplear palabras en uno o varios idiomas (Inbenta, 2022). Aunque todos los lenguajes naturales se basan en normas específicas, como la gramática y sintaxis, muchas veces contienen ambigüedades; el lenguaje natural también es creado de manera artificial, como

pueden ser, los lenguajes utilizados en programación informática (Inbenta, 2022).

Fases del lenguaje natural

1. **Tokenización:** Descompone el texto en pequeñas unidades semánticas u oraciones más simples.
2. **Etiquetar categorías gramaticales:** Marca cada palabra como verbo, adjetivo, sustantivo, pronombre, adverbio, entre otros.
3. **Eliminar palabras vacías:** Se filtran las palabras comunes que agregan poca o ninguna información, como, por ejemplo: los artículos o preposiciones.
4. **Derivar/lematizar:** Pasar las palabras o texto a su forma raíz y realizar la valoración del contexto del uso de las palabras.

Aprendizaje automático

El aprendizaje automático emplea un enfoque de manera estadística para resolver las consultas de los clientes o usuarios; esto significa que, la solución requiere de un entrenamiento con muchos datos e información, es decir, distintas formulaciones de solicitudes de usuarios, para que los algoritmos decidan de forma estadística, cómo responder a cada pregunta respectivamente (Oracle, 2022).

Se centra en la concordancia semántica, agregando concordancias procedentes del aprendizaje automático, permitiendo que el motor PLN analice y aprenda del comportamiento de los seres humanos, para mostrar los contenidos más populares en primer lugar (Oracle, 2022).

Aplicaciones del PLN

- Búsqueda avanzada de datos e información
- Chatbots
- Detección de entidades

- Detección de similitudes o anomalías en textos
- Clasificación automática de mensajes y documentos
- Anonimización de documentos
- Análisis de sentimientos y opiniones

Tecnología GPT

La tecnología GPT (Transformados preentrenado generativo) es un tipo de arquitectura de aprendizaje profundo que se llama transformador (EY Perú, 2023). Los mismos se basan en ideas de la autoatención, permitiendo que el modelo suponga la importancia de las distintas partes de la secuencia de entrada al realizar predicciones.

GPT emplea un tipo específico de arquitectura de transformador llamada decodificador, que se utiliza con grandes volúmenes de datos de texto sin supervisión; el proceso de preentrenamiento consiste en la predicción de las palabras que faltan en la secuencia de un texto determinada, en función del contexto brindado por las demás palabras.

Beneficios de GPT

- **Generación de lenguaje:** Los modelos de GPT pueden realizar la generación de lenguaje y texto, con el sonido natural que sea relevante y coherente para la entrada o el mensaje dado.
- **Versatilidad:** Los modelos de GPT se ajustan para una gran variedad de tareas de procesamiento de lenguaje natural, como la clasificación de texto, respuesta a preguntas y resúmenes.
- **Eficiencia:** Los modelos de GPT son muy eficaces y eficientes, siendo capaces de procesar grandes volúmenes de datos en un corto periodo de tiempo.
- **Adaptabilidad:** Los modelos de GPT pueden aprender de distintas fuentes de datos de texto, adaptándose a nuevos

contextos o dominios, lo cual los hace adaptables y versátiles a la amplia gama de aplicaciones.

- **Precisión:** Los modelos de GPT han demostrado la gran precisión en la variedad de tareas de PLN, incluyendo la clasificación de textos, traducción de los idiomas y respuesta a las preguntas.
- **Accesibilidad:** Los modelos de GPT tienen disponibilidad gratuita y pueden ser empleados por cualquier usuario con conexión a internet y que tenga conocimientos básicos de programación

GPT – 3: GPT – 3 es un nuevo modelo de inteligencia artificial que permite la generación de lenguaje natural o escrito, gracias a dicho algoritmo, el cliente o usuario solo debe escribir un párrafo y el sistema se encarga de completar el resto de texto de la manera más coherente posible (OpenAI, 2023).

Uso actual de GPT-3

El modelo GPT – 3 en la actualidad no es una solución verdaderamente práctica para solucionar diversos problemas industriales, por lo tanto, su utilización hoy en día se limita fundamentalmente a los investigadores, los cuales trabajan partiendo de una API.

El sistema todavía presenta muchas limitaciones y replicar el trabajo hecho por OpenAI es muy complejo, ya que, los costos computacionales son muy elevados (Santana, 2022). A esto, se le agregan los inconvenientes que tienen la mayoría de algoritmos de aprendizaje automático o más conocidos como sesgos.

Funcionamiento del procesamiento de lenguaje natural GPT-3

La tecnología GPT-3 está vinculada a una arquitectura de red neuronal denominada Transformers, que se especializa en el proceso y generación de texto automático permitiendo incurrir como una red neuronal especializada para las secuencias de texto,

llegando hasta un nivel capaz de aprender de forma automática a partir de la recolección de datos en conjuntos de alta calidad, este árbol de procesos naturales en medio del lenguaje, utiliza análisis morfológico, sintáctico y semántico para la comprensión del dialecto humano (Scheffler Zawadzki, 2022).

Con su sofisticada capacidad puede procesar enormes cantidades de texto y mediante el GPT-3 produce texto de mucha mejor calidad de manera automática para detectar el tipo de procesamiento del lenguaje llegando hasta usar un traductor de idiomas, producción de chatbots y asistentes virtuales, mejorando el análisis de sentimientos y la clasificación de texto (Scheffler Zawadzki, 2022).

Esta herramienta ha ayudado a los editores de contenido periodísticos, políticos y de edición a crear contenido de forma rápida y precisa. Dichas herramientas de IA pueden entrenarse para el reconocimiento de patrones generando contenido textual de manera eficiente. Esto significa que los editores ya no tienen que pasar horas escribiendo y editando contenido de manera manual, sino que pueden usar la tecnología GPT-3 para crear el contenido que ellos necesiten o requieran (Scheffler Zawadzki, 2022).

Ventajas de GPT-3

Muchos sectores se interesan mucho por usar esta tecnología, variando en el interés de la grande y pequeña empresa en cualquier sector que se quiera elaborar y las cuales se muestran reflejadas en las ventajas que se muestran a continuación (axolot, 2020):

- ***Avances en el procesamiento del lenguaje natural:*** Como tecnología, es un gran avance en NLP (procesamiento del lenguaje natural) porque permite un procesamiento del lenguaje más preciso y eficiente. Por lo tanto, proporciona una interacción más humana.

- ***Mayor eficiencia y productividad:*** Esto ahorra un tiempo valioso para los empleados de la empresa y aumenta su productividad.
- ***Mejorar la experiencia del usuario:*** Esto hace que la interacción con los servicios sea natural. De esta manera, crea una conexión fluida gracias a la respuesta de transmisión más humana.
- ***Mejorar las habilidades para la toma de decisiones:*** Gracias a los datos recopilados, puede ser una herramienta de gestión para la organización. De esta manera, proporciona ideas y sugerencias basadas en certezas más que en suposiciones.
- ***Minimización de errores:*** Primero, reducir el error humano a través de la inteligencia artificial. Las máquinas son más precisas y están menos sujetas a errores externos.
- ***Al es compatible con muchas industrias y sectores:*** Se puede utilizar en el campo de la salud o incluso en la industria de la aviación o el transporte.
- ***Al optimizado máximo:*** De esta forma, el dispositivo puede aprovechar al máximo su rendimiento y realizar múltiples tareas a la vez, haciendo a un lado las funciones innecesarias. Por ejemplo, los coches autónomos.

Desventajas de GPT-3

Como bien se apreciaron los beneficios de usar GPT-3, también tiene sus limitaciones que se consideran para el uso intra informático y que podrían afectar dependiendo de su mal uso, tales como (wearecontent, 2023):

- ***Sesgo y falta de variedad:*** Al usar los datos en la capacitación, puede reflejar el sesgo y la falta de variedad en dichos datos. Esto puede conllevar a resultados discriminatorios y perpetuar los estereotipos.

- **Privacidad y seguridad:** Al igual que con cualquier diseño de IA, existen problemas con el almacenamiento y el uso de datos.
- **Confíe en un modelo:** Confiar en un solo modelo de IA puede colapsar. Por lo tanto, si es necesario, será difícil encontrar una solución alternativa.
- **La disponibilidad es limitada:** La alta demanda del servicio puede dar como resultado una cobertura insuficiente y, en última instancia, la falta de disponibilidad para las partes interesadas.
- **Las palabras se repiten:** Aunque Smart se enorgullece de generar contenido original, en algunas aplicaciones puede cometer el error de SEO de copiar palabras en exceso.
- **Creciente desempleo:** Recordemos que, si las máquinas pueden realizar las mismas tareas que los humanos, provocarán un aumento del desempleo.
- **Cuesta mucho:** Invertir en IA requiere un gran compromiso, por lo que las empresas más grandes son las primeras en implementarla en sus procesos y gestión.
- **Carece de sentimientos y valores:** A pesar de lo eficiente que es la tecnología, él no es humano y no tiene sentimientos. Entonces, como mencionamos anteriormente, no tiene fronteras ni barreras morales. La situación puede ser muy peligrosa si no se detiene (Wearecontent, 2023).

Metodología

La presente investigación tiene un enfoque cuantitativo con el cual se plantea comprender las relaciones de causa y efecto de los procesamiento de aprendizaje automático del lenguaje natural con tecnología GPT, se caracterizó por ser una investigación básica permitiendo ampliar la comprensión. La finalidad de la investigación es exploratoria que permitió identificar y recopilar la infor-

mación pertinente logrando definir y aportar al estudio de prototipos de procesamiento de lenguaje natural con tecnología GPT.

Métodos

Revisión Bibliográfica: Permitted recopilar y analizar información relevante sobre el tema de investigación, a partir de la revisión de libros, fuentes de artículos, revistas, sitios web, entre otros.

Inducción – Deducción: Este método se empleó a partir de la hipótesis, la cual es puesta a comprobación.

Histórico Lógico: Se utilizó en la búsqueda de los antecedentes investigativos relacionados al procesamiento del lenguaje natural.

Análisis Síntesis: Este método se empleó en la revisión y el análisis que permitió establecer la síntesis del modelo de técnicas de aprendizaje automático.

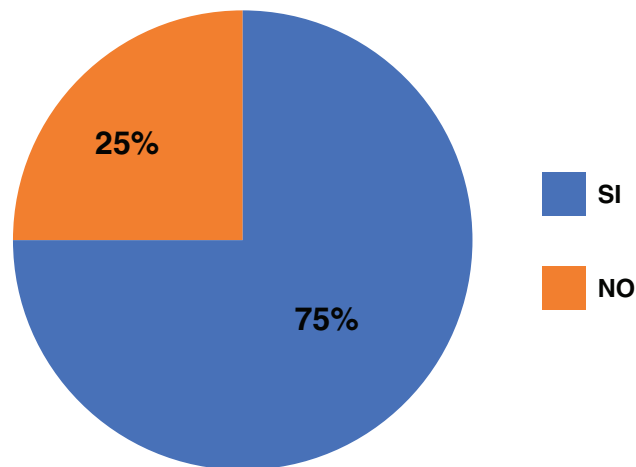
Encuesta: se le realizó a los docentes de la carrera tecnología de la información para conocer su criterio o su opinión referente al procesamiento de lenguaje natural bajo las técnicas de aprendizaje automático

Resultados

La población que se tomó en consideración son 36 docentes de la carrera de Tecnologías de la Información de la Universidad Estatal del Sur de Manabí.

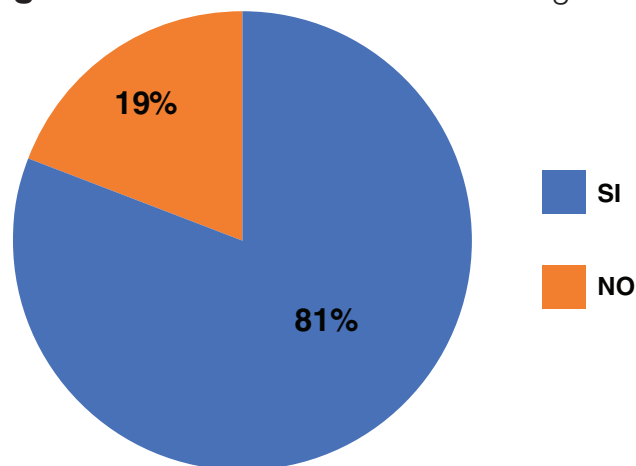
1. ¿Ha escuchado acerca de la inteligencia artificial (AI)?

Figura 1. Conocimiento sobre Inteligencia Artificial
¿Ha escuchado acerca de la inteligencia artificial (AI)?



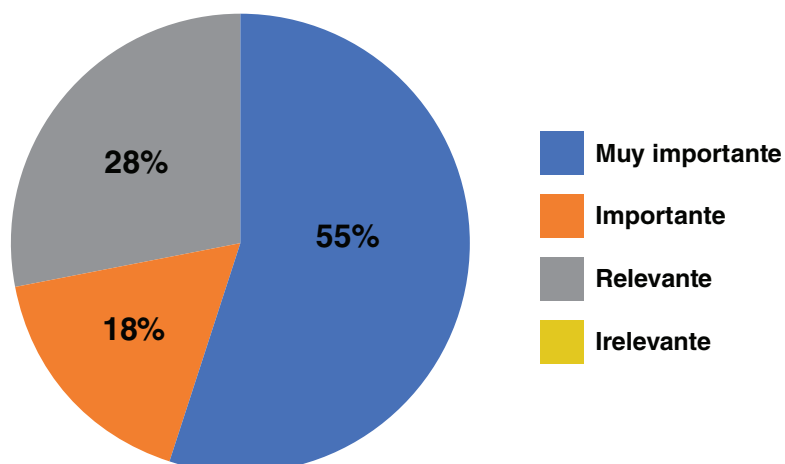
2. ¿Conoce usted sobre la tecnología GPT?

Figura 2. Conocimiento sobre Tecnología GPT



3. ¿Desde su perspectiva, ¿Qué importancia tiene la utilización de la tecnología GPT, como apoyo al proceso de enseñanza en su área?

Figura 3. Importancia de utilizar GPT



Los resultados en base a la encuesta realizada permitieron centrar la atención en la necesidad del uso de las tecnologías de procesamiento de lenguaje natural basadas en técnicas de aprendizaje automático y tecnologías GPT, desarrollando el potencial en la resolución de problemas lingüísticos complejos, conduciendo a la creación de solución innovadora.

Conclusiones

- La combinación de tecnología GPT y técnicas de aprendizaje automático ha allanado el camino para una mayor accesibilidad y democratización del Procesamiento de Lenguaje Natural. Estos prototipos avanzados han permitido que una variedad de desarrolladores y empresas integren capacidades de PLN en sus aplicaciones y servicios, lo que antes era un privilegio de pocos. Esta democratización ha abierto nuevas oportunidades para innovar y mejorar la experiencia del usuario, fomentando así un ecosistema más inclusivo y diverso en el campo del PLN.
- La adaptabilidad de los prototipos GPT es un factor clave en su éxito. Gracias al aprendizaje automático, estos modelos pueden abordar eficazmente la diversidad de idiomas y dialectos, permitiendo una comunicación más natural y enriquecedora a nivel global. Sin embargo, a pesar de los avances logrados, aún existen desafíos persistentes, como la comprensión contextual y la detección de sesgos, que requieren una atención continua para mejorar la precisión y equidad de estos sistemas.
- El futuro del Procesamiento de Lenguaje Natural con tecnología GPT y aprendizaje automático es prometedor. A medida que la investigación y el desarrollo continúan, se espera que estos prototipos evolucionen aún más, superando las limitaciones actuales y abriendo nuevas fronteras en la inteligencia artificial. Con el tiempo, es posible que veamos avan-

ces en la generación de lenguaje más creativo y en la comprensión profunda de la semántica y el contexto, lo que podría revolucionar aún más nuestra forma de comunicarnos e interactuar con la tecnología.

Bibliografía

- Axolot. (10 de 10 de 2020). ¿Qué es y para que sirve el algoritmo GPT-3 de OpenAI? Un algoritmo que permite generar contenido automáticamente. Obtenido de ¿Qué es y para que sirve el algoritmo GPT-3 de OpenAI? Un algoritmo que permite generar contenido automáticamente: <https://axolotagencia.com/que-es-y-para-que-sirve-gpt-3-de-openai/#:~:text=Una%20de%20las%20ventajas%20m%C3%A1s,centrarse%20en%20tarear%20m%C3%A1s%20importantes>.
- EY Perú. (2023). ey.com. Obtenido de [https://www.ey.com/es_pe/revista-execution/tendencias/gpt-beneficios-tecnologia-futuro-modelo-lenguaje#:~:text=La%20tecnologia%20C3%ADa%20detr%C3%A1s%20de%20GPT,del%20lenguaje%20natural%20\(NLP\)](https://www.ey.com/es_pe/revista-execution/tendencias/gpt-beneficios-tecnologia-futuro-modelo-lenguaje#:~:text=La%20tecnologia%20C3%ADa%20detr%C3%A1s%20de%20GPT,del%20lenguaje%20natural%20(NLP))
- Inbenta. (2022). inbenta.com. Obtenido de <https://www.inbenta.com/es/tecnologia/tecnologia-en-lenguaje-natural/>
- Maldonado Zuñiga, K., Lucas Delgado, H., Vera Velázquez, R., & Cables Fernández, E. (2023). La inteligencia artificial y su impacto en la educación superior. Serie Científica De La Universidad De Las Ciencias Informáticas, 16(7), 204-215. Recuperado a partir de <https://publicaciones.uci.cu/index.php/serie/article/view/1378>
- OpenAI. (2023). openai.com. Obtenido de <https://openai.com/product>
- Oracle. (2022). oracle.com. Obtenido de <https://www.oracle.com/pe/artificial-intelligence/what-is-natural-language-processing/>
- Santana, C. (2022). bbvaopenmind.com. Obtenido de <https://www.bbvaopenmind.com/tecnologia/inteligencia-artificial/gpt-3-sistemas-de-escritura-ia-calidad-contenidos-online/>
- Sayago, S. (2023). mpiua.invid.udl.cat. Obtenido de <https://mpiua.invid.udl.cat/fases-mpiua/prototipado/que-es-un-prototipo/>
- Scheffler Zawadzki, E. (8 de 12 de 2022). ¿Qué es GPT-3? Le pedimos al modelo de inteligencia artificial que se definiera a sí mismo y esto es lo que nos respondió. (entrepreneur) Recuperado el 10 de 3 de 2023, de <https://www.entrepreneur>.

com/es/tecnologia/que-es-gpt-3-le-pedimos-al-modelo-de-inteligencia/440699

Wearecontent. (8 de 03 de 2023). GPT-3 y chatGPT: el futuro de la inteligencia artificial. (wearecontent) Recuperado el 10 de 03 de 2023, de <https://www.wearecontent.com/blog/news/gpt-3>

Cómo citar: Ordóñez Palma, M. J., & Gutiérrez García, J. L. (2023). Prototipos de procesamiento de lenguaje natural bajo el modelo de técnicas de aprendizaje automático. *Journal TechInnovation*, 2(2), 34–43. Recuperado a partir de <https://revistas.unesum.edu.ec/JTI/index.php/JTI/article/view/49>