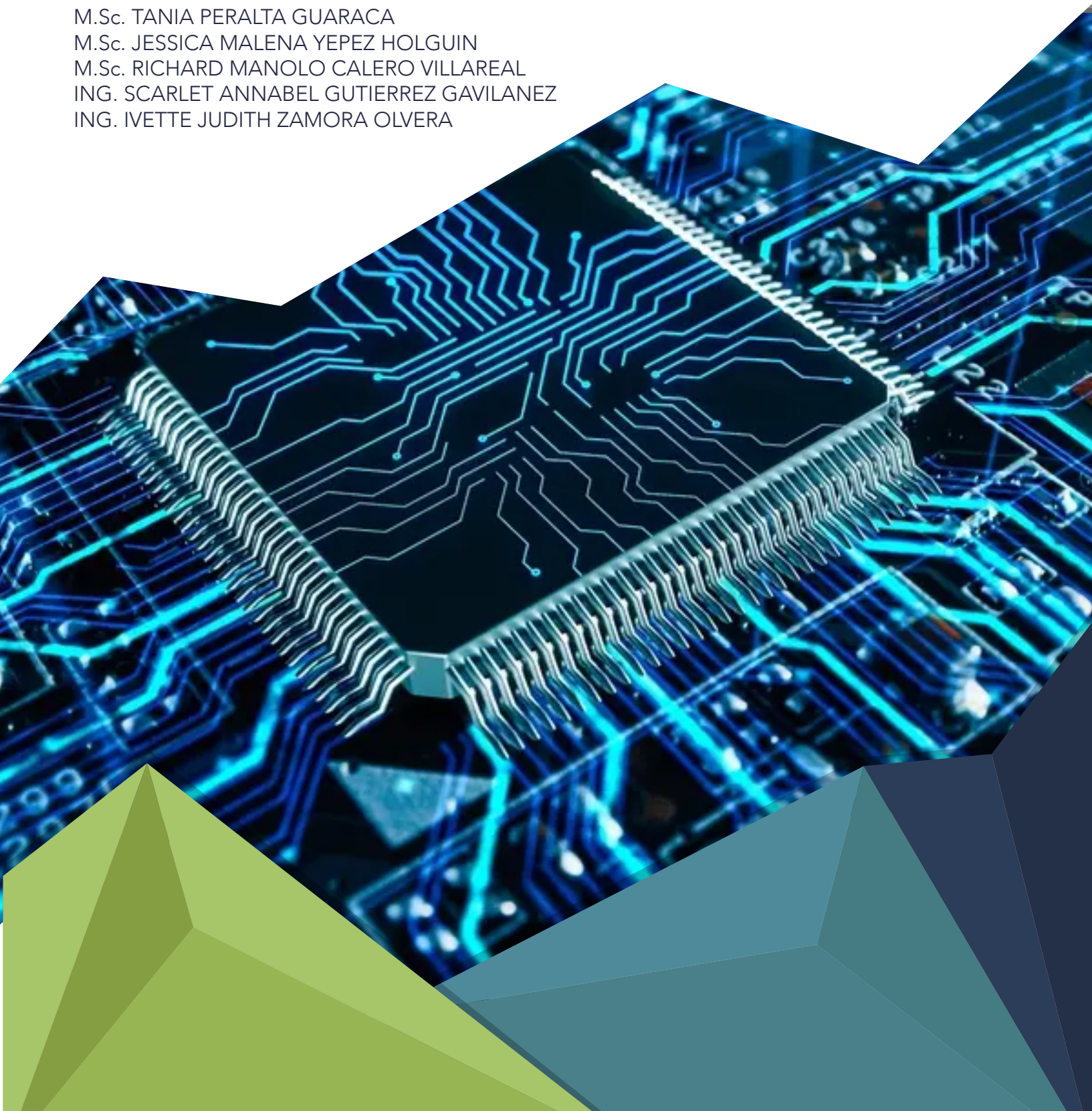


Evaluación del desempeño del modelo generative Pre-trained transformer gpt-4 y sus sucesores en la predicción de la complejidad léxica en textos públicos

Ph.D. JENNY ORTIZ ZAMBRANO
M.Sc. MIRELLA CARMINA ORTIZ ZAMBRANO
M.Sc. TANIA PERALTA GUARACA
M.Sc. JESSICA MALENA YEPEZ HOLGUIN
M.Sc. RICHARD MANOLO CALERO VILLAREAL
ING. SCARLET ANNABEL GUTIERREZ GAVILANEZ
ING. IVETTE JUDITH ZAMORA OLVERA





© **Ph.D. JENNY ORTIZ ZAMBRANO**

Docentes de la Universidad de Guayaquil

M.Sc. MIRELLA CARMINA ORTIZ ZAMBRANO

Docentes de la Universidad de Guayaquil

M.Sc. TANIA PERALTA GUARACA

Docentes de la Universidad de Guayaquil

M.Sc. JESSICA MALENA YEPEZ HOLGUIN

Docentes de la Universidad de Guayaquil

M.Sc. RICHARD MANOLO CALERO VILLAREAL

Docentes de la Universidad de Guayaquil

ING. SCARLET ANNABEL GUTIERREZ GAVILANEZ

ING. IVETTE JUDITH ZAMORA OLVERA

© Editorial Grupo Compás, 2025

Guayaquil, Ecuador

www.grupocompas.com

<http://repositorio.grupocompas.com>

Primera edición, 2025

ISBN: 978-9942-53-098-1

Distribución online

 Acceso abierto

Cita

Ortiz, J., Ortiz, M., Peralta, T., Yopez, J., Calero, R., Gutierrez, S., Zamora, I. (2025) Evaluación del desempeño del modelo generative Pre-trained transformer gpt-4 y sus sucesores en la predicción de la complejidad léxica en textos públicos. Editorial Grupo Compás

Este libro ha sido debidamente examinado y valorado en la modalidad doble par ciego con fin de garantizar la calidad de la publicación. El copyright estimula la creatividad, defiende la diversidad en el ámbito de las ideas y el conocimiento, promueve la libre expresión y favorece una cultura viva. Quedan rigurosamente prohibidas, bajo las sanciones en las leyes, la producción o almacenamiento total o parcial de la presente publicación, incluyendo el diseño de la portada, así como la transmisión de la misma por cualquiera de sus medios, tanto si es electrónico, como químico, mecánico, óptico, de grabación o bien de fotocopia, sin la autorización de los titulares del copyright.

ÍNDICE GENERAL	
ÍNDICE DE TABLAS	11
ÍNDICE DE FIGURAS	12
ÍNDICE DE ECUACIONES	13
ABREVIATURAS	14
SIMBOLOGÍA	15
RESUMEN.....	16
ABSTRACT.....	17
INTRODUCCIÓN	18
CAPÍTULO I	22
PLANTEAMIENTO DEL PROBLEMA	22
Descripción de la situación problemática	22
Ubicación del problema en un contexto.....	22
Situación conflicto nudos críticos.....	23
Delimitación del problema	22
Evaluación del Problema.....	22
Causas y consecuencias del problema	24
Formulación del problema	25
Objetivos del proyecto	25
Objetivo general	25
Objetivos específicos.....	25
Alcance del problema.....	25
Justificación e importancia.....	26
Limitaciones del estudio.....	27
CAPÍTULO II	30
MARCO TEÓRICO	30
Antecedentes del estudio.....	30
Fundamentación teórica.....	33
Complejidad textual en documentos públicos.....	34
Factores que influyen en la complejidad textual.....	35
Impacto de la complejidad textual en la participación ciudadana	35

Métodos de evaluación de complejidad textual	36
Aplicación de modelos Transformer en la evaluación de complejidad textual	37
Aplicaciones de simplificación textual en documentos.....	37
La simplificación del lenguaje administrativo en la mejora de la accesibilidad y comprensión de documentos públicos.....	38
GPT-4: Características y aplicaciones	38
GPT-4 y su aplicación en el análisis de complejidad textual.....	39
Métodos tradicionales de evaluación de complejidad textual	40
Comparativa de modelos basados en la arquitectura Transformer	41
GPT-4 y su potencial en el análisis de documentos públicos	42
GPT-4-Turbo para la predicción de la complejidad léxica.....	42
GPT-4o para la predicción de la complejidad léxica	43
Zero-Shot y Few-Shot Learning.....	43
Zero-Shot Learning en Modelos de Lenguaje: Enfoque en GPT-4 y sus Sucesores.....	44
Few-Shot Learning en Modelos de Lenguaje: Enfoque en GPT-4 y sus Sucesores.....	44
Ética y consideraciones en el uso de modelos generativos para documentos públicos	45
Variables de investigación e hipótesis	46
Revisión sistemática	46
Planteamiento de la pregunta estructurada	46
Búsqueda en bases de datos	46
Selección de artículos.....	47
Extracción de datos.....	47
Meta-análisis	47
Preguntas científicas a contestarse	31
Definiciones conceptuales.....	31
Tipo de investigación	31
Diseño metodológico de la investigación	49
Enfoque de la Investigación	49
Población	49
Características lingüísticas	52
Grupos experimentales.....	64

Tratamiento (Modificación).....	64
Test Estadísticos	66
Beneficiarios directos e indirectos del proyecto	69
Entregables del proyecto	70
Propuesta.....	70
Criterios de validación de la propuesta.....	71
Resultados	73
Características lingüísticas	73
Aplicación del modelo GPT-4 y sus sucesores GPT-4 Turbo y GPT-4o.....	74
Evaluación de la predicción de la complejidad léxica entre modelos.....	80
CAPÍTULO IV	95
CONCLUSIONES Y RECOMENDACIONES	95
Conclusiones.....	95
Recomendaciones	97
Trabajos futuros.....	97
REFERENCIAS BIBLIOGRÁFICAS	99

ÍNDICE DE TABLAS

Tabla

1 Delimitación del problema.....	22
Tabla 2 Matriz de causas y consecuencias del problema	24
Tabla 3 Corpus GovAIEc	51
Tabla 4	52
Tabla 5 Fórmulas de las métricas de legibilidad	60
Tabla 6 Clasificación de dificultad de lectura - Flesch Reading Ease..	61
Tabla 7 Clasificación de dificultad de lectura - RIX Score.....	62
Tabla 8 Clasificación de dificultad de lectura - Gunning-Fog Index ...	62
Tabla 9 Clasificación de dificultad de lectura - SMOG Score.....	63
Tabla 10 Interpretación de Coeficientes de Correlación	68
Tabla 11 Expertos	71
Tabla 12 Criterios de evaluación	71
Tabla 13 Juicio de expertos	72
Tabla 14 Características lingüísticas	73
Tabla 15 Palabras complejas GPT-4 / Zero-Shot Learning	75
Tabla 16 Palabras complejas GPT-4 Turbo / Zero-Shot Learning	76
Tabla 17 Palabras complejas GPT-4o / Zero-Shot Learning.....	76
Tabla 18 Palabras complejas GPT-4 / Few-Shot Learning.....	77
Tabla 19 Palabras complejas GPT-4 Turbo / Few-Shot Learning	78
Tabla 20 Palabras complejas GPT-4o / Few-Shot Learning	78
Tabla 21 Resultado de los sucesores	79

ÍNDICE DE FIGURAS

Figura 1 Función de Normalización.....	63
Figura 2 Prompt Zero-Shot Learning.....	65
Figura 3 Prompt Few-Shot Learning.....	65
Figura 4 Cálculo de Métricas de Evaluación para Predicción de Complejidad	69
Figura 5 Porcentaje de coincidencias de los modelos	92
Figura 6 Rendimiento de los modelos en función del MAE.....	93

ÍNDICE DE ECUACIONES

Ecuación 1 Índice de legibilidad de Flesch.....	60
Ecuación 2 Índice RIX	60
Ecuación 3 Índice de Gunning-Fog	60
Ecuación 4 Índice de SMOG.....	60
Ecuación 5 Mean Absolute Error	66
Ecuación 6 Mean Squared Error	67
Ecuación 7 Root Mean Squared Error	67
Ecuación 8 R-squared.....	67
Ecuación 9 Pearson Correlation Coefficient.....	67
Ecuación 10 Spearman Correlation Coefficient.....	68
Ecuación 11 Diferencia porcentual	69

ABREVIATURAS

CC.MM.FF	Facultad de Ciencias Matemáticas y Físicas
Ing.	Ingeniero
LSI	Licenciatura en Sistemas Informáticos
M.Sc.	Máster
UG	Universidad de Guayaquil
AI	Artificial Intelligence (Inteligencia Artificial)
GPT	Generative Pre-trained Transformer
GPT-4	Generative Pre-trained Transformer 4
LPN	Natural Language Processing
PCL	Prediction of Lexical Complexity
ZSL	Zero-Shot Learning (Aprendizaje de cero ejemplos)
FSL	Few-Shot Learning (Aprendizaje con pocos ejemplos)
API	Application Programming Interface
MAE	Mean Absolute Error (Error Absoluto Medio)
MSE	Mean Squared Error (Error Cuadrático Medio)
RMSE	Root Mean Squared Error (Raíz del Error Cuadrático Medio)
R ²	Coefficiente de Determinación (R-cuadrado)
SMOG	Simple Measure of Gobbledygook
CNT	Corporación Nacional de Telecomunicación
SRI	Servicio de Rentas Internas
MING	Municipio de Guayaquil
ATM	Autoridad de Tránsito Municipal de Guayaquil
CNE	Consejo Nacional Electoral

SIMBOLOGÍA

m	Número de observaciones
Σ	Sumatoria
ρ	Coefficiente de correlación de Pearson
r	Coefficiente de correlación de Spearman
$=$	Igual
\approx	Aproximadamente igual
$\%$	Porcentaje

RESUMEN

El proyecto tiene como objetivos calcular nuevas características lingüísticas, aplicar los modelos Generative Pre-trained Transformer GPT4 y sus sucesores en el conjunto de datos GovAIEc mediante la utilización de las técnicas zero-shot learning y few-shot learning y por último evaluar la complejidad de los textos de los documentos públicos mediante el análisis de los resultados obtenidos de las ejecuciones de los modelos. La orientación teórica se fundamenta en el análisis de modelos de lenguaje avanzado y su capacidad para comprender y procesar estructuras lingüísticas complejas, apoyándose en estudios recientes sobre procesamiento de lenguaje natural (NLP) y métricas de evaluación de complejidad léxica. En cuanto a la metodología de la investigación es cuantitativa y descriptiva, con un enfoque cuasi experimental, en donde se aplican los modelos GPT-4, GPT-4o y GPT-4 Turbo, también se describen las características lingüísticas y se hace mención a los test estadísticos que se aplicaran a las ejecuciones de los modelos. Entre los hallazgos durante la investigación están las características lingüísticas que dan un total de 40, entre ellas están el número de vocales, numero de morfemas, numero de sustantivos, las cuales permiten analizar la complejidad léxica del corpus, que es donde se aplicó el cálculo de las características halladas, en cuanto a los modelos para las ejecuciones se utilizaron las técnicas de few-shot y zero-shot para cada uno modificando el prompt en cada técnica. En conclusión, el rendimiento con mayor precisión en función del MAE lo tiene el modelo GPT-4 con few-shot (0.24641), mientras que el rendimiento con menos precisión lo tiene el modelo GPT-4o con few-shot (0.3889), a rasgos generales los tres primeros rendimientos con el MAE más bajo son con la técnica few-shot learning, lo que sugiere que esta técnica es más efectiva para mejorar la precisión de los modelos que la técnica de zero-shot learning.

Palabras clave: Gpt-4-Turbo, Gpt-4-o, Zero-shot learning, few-shot learning, predicción de la complejidad lexica.

ABSTRACT

The project aims to calculate new linguistic features, apply the Generative Pre-trained Transformer GPT4 and its successors models on the GovAlEc dataset using zero-shot learning and few-shot learning techniques, and finally evaluate the complexity of public document texts by analyzing the results obtained from the model executions. The theoretical orientation is based on the analysis of advanced language models and their ability to understand and process complex linguistic structures, supported by recent studies on natural language processing (NLP) and lexical complexity assessment metrics. The research methodology is quantitative and descriptive, with a quasi-experimental approach, where the GPT-4, GPT-4o and GPT-4 Turbo models are applied. The linguistic characteristics are also described and the statistical tests that will be applied to the model executions are mentioned. Among the findings during the research are the linguistic characteristics that give a total of 40, among them are the number of vowels, number of morphemes, number of nouns, which allow to analyze the lexical complexity of the corpus, which is where the calculation of the characteristics found was applied, as for the models for the executions, the few-shot and zero-shot techniques were used for each one, modifying the prompt in each technique. In conclusion, the performance with the highest accuracy as a function of MAE is the GPT-4 model with few-shot (0.24641), while the performance with the lowest accuracy is the GPT-4o model with few-shot (0.3889), overall the first three performances with the lowest MAE are with the few-shot learning technique, suggesting that this technique is more effective in improving the accuracy of the models than the zero-shot learning technique.

Key words: Gpt-4-Turbo, Gpt-4-o, Zero-shot learning, few-shot learning, lexical complexity prediction.

INTRODUCCIÓN

La comprensión de los textos en documentos públicos es un componente esencial para la correcta ejecución de la participación ciudadana en sociedades democráticas, así como la transparencia de la información, según (Poblete & González, 2018) transparencia no se limita a la publicación de la información, sino que también implica asegurarse que sea comprendida por las personas, ya que si un texto no logra comunicar su significado pierde su función comunicativa.

En esta investigación se utiliza un conjunto de datos conformados por textos oficiales de las instituciones gubernamentales de la ciudad de Guayaquil, como el Consejo Nacional Electoral (CNE), el Municipio de Guayaquil (MIMG), el Servicio de Rentas Internas (SRI), la Corporación Nacional de Telecomunicación (CNT) y la Autoridad de Tránsito Municipal de Guayaquil (ATM), ya que contienen información crucial que debe ser accesible para todos los ciudadanos. Sin embargo, el lenguaje técnico y burocrático que caracteriza a muchos de estos documentos a menudo dificulta su comprensión, excluyendo a sectores significativos de la población del acceso pleno a sus derechos y deberes.

De esta manera es necesario explorar herramientas que puedan evaluar la complejidad del lenguaje de estos documentos, para esto se plantea evaluar la predicción de la complejidad léxica con modelos de GPT-4, los cuales según un estudio realizado por (Cesteros, 2023) indican que presentan un enfoque prometedor para la simplificación léxica, especialmente en el contexto de palabras complejas previamente identificadas, donde el modelo muestra un rendimiento elevado a pesar de utilizar datos de entrenamiento etiquetados de forma muy limitada. El objetivo de este trabajo de investigación es evaluar el desempeño del modelo Generative Pre-trained Transformer GPT-4 y sus sucesores en la predicción de la complejidad léxica en textos públicos en español, a través de un análisis que determine el rendimiento de los modelos basados en arquitectura Transformer.

Para una mejor comprensión acerca del tema central de la investigación, este documento presenta la siguiente estructura:

En el capítulo I, se plantea la problemática donde se describe la situación con respecto a la complejidad léxica en documentos públicos y el uso de los modelos de GPT, se delimita, evalúa y formula el problema, así como también se plantean los objetivos del proyecto y su alcance.

En el capítulo II, se dan a conocer los antecedentes del estudio, así como su fundamentación teórica basada en una exhaustiva revisión de artículos relacionados al tema, además se presentan las variables, hipótesis, preguntas científicas a contestarse y las definiciones conceptuales.

En el capítulo III, se expone la metodología a seguir para la investigación, en donde se indica el diseño metodológico de la investigación, así como los beneficiarios del proyecto, la propuesta, los criterios de validación y los resultados.

Finalmente, en el capítulo IV, se presentan las conclusiones de la investigación, recomendaciones y se hace mención de futuras líneas de investigación, es decir a futuros trabajos.

CAPÍTULO I

PLANTEAMIENTO DEL PROBLEMA

Descripción de la situación problemática

Ubicación del problema en un contexto

Los documentos públicos son mayormente impartidos por el estado nacional de cada país. Según (Meneses Pacheco, 2018), estos documentos se caracterizan por la presencia de un conjunto de disposiciones legales que regulan procesos, es decir, comunican hechos importantes para el desarrollo del país y su ciudadanía, por lo que es de vital importancia la comprensión de estos por parte de cada individuo miembro de la sociedad.

Para el autor (Fuentes-Monsálves, 2009) la competencia y comprensión lectora de cualquier material escrito "se entienden hoy como un prerrequisito para el aprendizaje y el desenvolvimiento de cualquier tarea cotidiana". En este sentido, la comprensión de los textos en documentos públicos resulta esencial para garantizar el cumplimiento de las disposiciones legales contenidas en ellos.

Asimismo, la autora (Ruohonen, 2021) recalca que "la comprensión de diferentes textos es una exigencia de la vida actual. Independientemente de su origen, los ciudadanos necesitan comprender muchos tipos de textos para participar en una sociedad y gestionar sus vidas".

Por consiguiente, la misma autora expresa que la comprensión de las leyes que mayormente contienen documentos públicos por parte de los ciudadanos sigue siendo un desafío crítico especialmente para quienes tienen niveles bajos de alfabetización, ya que incluso los sitios de asesoramiento jurídico para muchos ciudadanos suelen ser inaccesibles. Esta indica que el problema se divide en dos razones:

1. *Confusión intencional*: A menudo empresas y representantes legales suelen redactar documentos complejos para evitar cumplir con requisitos de transparencia.
2. *Complejidad innecesaria*: Profesionales como abogados, académicos y legisladores tienden a emplear un lenguaje complicado debido a la complejidad del tema, lo que dificulta la comprensión de los textos.

Por esto, la autora expresa que desde hace tiempo se reconoce la importancia de un lenguaje claro y sencillo, tanto en la administración pública como en la redacción de leyes, recomendándose mejoras lingüísticas para favorecer la transparencia y accesibilidad.

Situación conflicto nudos críticos

Como se menciona anteriormente, el problema surge por el lenguaje complejo con el que se redactan estos documentos, afectando especialmente a personas con bajos niveles de alfabetización. En Ecuador, según el último censo realizado en el año 2022 del Instituto Nacional de Estadística y Censos (INEC) hay 472.228 personas en condición de analfabetismo (El Universo, 2024) , lo que ocasiona una disminución de la comprensión de textos, siendo así que no se cumple el derecho de los ciudadanos a tener información clara de manera que se posibilite la participación ciudadana en la toma de decisiones de interés general según lo menciona el artículo 4 de la LOTAIP (Ley Orgánica de Transparencia y Acceso a la Información Pública). LOTAIP (2004)

.

Delimitación del problema

Para abordar el problema, la investigación se centra en el campo de la inteligencia artificial y el procesamiento de lenguaje natural de los diferentes modelos generativos que se han desarrollado a lo largo del tiempo. Dentro de este campo, el área de interés son los modelos GPT-4 y sus sucesores, específicamente en la predicción de la complejidad léxica en textos públicos, así también el aspecto principal del estudio es la identificación de las limitaciones y capacidades de estos modelos para facilitar la comprensión de los textos por parte de la ciudadanía, con el objetivo de contribuir a una mayor claridad. El tema de este proyecto se titula: "Evaluación del desempeño del Modelo Generative Pre-Trained Transformer GPT-4 y sus sucesores en la predicción de la complejidad léxica en textos públicos".

Tabla

1

Delimitación del problema

Delimitador	Descripción
Campo	Inteligencia artificial y el procesamiento de lenguaje natural
Área	Modelos GPT-4 y sus sucesores
Aspecto	Identificación de las limitaciones y capacidades de estos modelos para facilitar la comprensión de los textos públicos
Tema	Evaluación del Desempeño del Modelo Generative Pre-Trained Transformer GPT-4 y sus sucesores en la predicción de la complejidad léxica en textos públicos

Elaborado por: Scarlet Gutierrez e Ivette Zamora.

Evaluación del Problema

El eje principal del problema a estudiar cómo se indicó en puntos anteriores es la complejidad léxica en textos públicos mediante el uso de modelos GPT-4 y sus sucesores, para este caso en particular, se ha contemplado utilizar los sucesores GPT-4 Turbo y GPT-4o.

Los documentos a tomarse en cuenta para el estudio serán de entidades públicas como: El Consejo Nacional Electoral (CNE), el Municipio de Guayaquil (MIMG), el Servicio de Rentas Internas (SRI) , la Corporación Nacional de Telecomunicación (CNT) y la

Autoridad de Tránsito Municipal de Guayaquil (ATM), por lo que son documentos de mayor relevancia en la población de la ciudad de Guayaquil y hacen que su comprensión sea necesaria para que los ciudadanos lleven a cabo sus derechos o participen de los comunicados que se realizan en estos documentos.

Los resultados esperados del estudio serán útiles para trabajos futuros acerca del rendimiento de los modelos de GPT-4 en relación con la predicción de la complejidad léxica, además se busca contribuir a la claridad de los documentos, es decir al léxico con que se redactan dichos documentos siendo así que las personas más beneficiadas serán los residentes de la ciudad de Guayaquil, lo que también contribuirá a generar mayor confianza con las instituciones seleccionadas.

Causas y consecuencias del problema

Tabla

2

Matriz de causas y consecuencias del problema

Causas	Consecuencias
C1. Textos no adaptados a diferentes niveles educativos.	E1. Excluye a personas con menor formación educativa, afectando su acceso a la información.
C2. Escasa claridad en la redacción de textos gubernamentales.	E2. Confusión y malentendidos entre los ciudadanos, afectando así el ejercicio de derechos.
C3. Términos legales y técnicos excesivos.	E3. Confusión y desinformación entre los ciudadanos, lo que resulta en una comprensión errónea de sus derechos y obligaciones.
C4. Falta de ejemplos prácticos en los documentos.	E4. Desorientación al ejecutar la información, lo que lleva a errores en los procedimientos y frustración con el sistema gubernamental.
C5. Excesiva longitud de los documentos.	E5. Pérdida del interés o no leen completamente los documentos, causando desinformación.
C6. Inexistencia de variantes simplificadas de documentos complejos.	E6. La falta de estas variantes puede llevar a que los ciudadanos no comprendan la información y resulte en malentendidos.

Nota: Esta tabla refleja el análisis causal que se realizó en base a la recopilación inicial de información de la situación problemática que genera el proyecto mediante la aplicación de la Metodología

de Marco Lógico, se considera por ello datos relevantes de la fase de investigación del proyecto. **Elaborado por:** Scarlet Gutierrez e Ivette Zamora.

Formulación del problema

¿El rendimiento del modelo Generative Pre-trained Transformer GPT-4 y sus sucesores en la predicción de la complejidad léxica de los textos públicos demuestran un desempeño notable en comparación con otros modelos basados en la arquitectura Transformer?

Objetivos del proyecto

Objetivo general

Evaluar el desempeño del modelo Generative Pre-trained Transformer GPT-4 y sus sucesores en la predicción de la complejidad léxica en textos públicos en español, a través de un análisis que determine el rendimiento de los modelos basados en arquitectura Transformer.

Objetivos específicos

1. Calcular nuevas características lingüísticas a nivel de palabra y a nivel de oración para fortalecer el conjunto de datos GovAIEc.
2. Aplicar los Modelo Generative Pre-trained Transformer GPT4 y sus sucesores (GPT-4 Turbo y GPT-4o) en el conjunto de datos GovAIEc mediante la utilización de las técnicas Zero-Shot learning y Few-shot learning.
3. Análisis comparativo del rendimiento de los modelos GPT-4 y sus sucesores en la predicción de la complejidad léxica.

Alcance del problema

El presente proyecto se centra en la evaluación del rendimiento del Modelo Generative Pre-trained Transformer GPT-4 en relación con la complejidad léxica de los textos públicos emitidos por las entidades gubernamentales; CNT, SRI, CNE, MIMG y ATM. Este análisis se llevará a cabo a través de un enfoque comparativo que incluirá el modelo GPT-4 y sus sucesores (GPT-4 Turbo y GPT-4o).

El estudio se limitará a un corpus de textos basados en las notificaciones e instrucciones de trámites burocráticos seleccionados de las entidades públicas anteriormente mencionadas, centrándose en su complejidad léxica y

características lingüísticas con el fin de identificar las barreras en la comprensión lectora a la que se enfrentan los usuarios, tal como lo confirman los resultados de investigaciones previas (Lyons III, 2018) muestran que, efectivamente, los documentos gubernamentales son difíciles de entender para los usuarios de todos los niveles educativos.

Se calcularán características lingüísticas específicas relacionadas con la complejidad de las palabras y otros indicadores que contribuyan a la evaluación de la complejidad de los textos de los documentos públicos, con el fin de identificar patrones de complejidad y su impacto en la comprensión lectora.

Se evaluará el rendimiento del Modelo Generative Pre-trained Transformer GPT-4 y sus sucesores en la comprensión e identificación de las palabras complejas de los textos de documentos públicos, utilizando la técnica de zero-shot learning y few-shot learning, como también lo demostraron (Mo & Hu, 2024), quienes indicaron que modelos zero-shot como ChatGPT y Llama-2 pueden modular la complejidad del texto.

El alcance del estudio se limitará a los documentos públicos de las entidades anteriormente mencionadas y a los sucesores del modelo GPT-4 indicadas y disponibles hasta la fecha de esta investigación. Asimismo, no se incluirán documentos de otras entidades gubernamentales ni se analizarán otros modelos de aprendizaje profundo en este estudio.

Justificación e importancia

La investigación sobre la evaluación de la complejidad léxica de los textos de documentos públicos utilizando el Modelo Generative Pre-trained Transformer GPT-4 y sus sucesores se justifica por varias razones fundamentales:

En primera instancia, esta investigación es importante para mejorar la accesibilidad de la información gubernamental. Al evaluar y simplificar los textos complejos, se facilita la comprensión de estos por parte de ciudadanos con diferentes niveles educativos y conocimientos técnicos, lo cual es importante para fomentar su participación en procesos administrativos y el ejercicio de sus derechos.

La investigación tiene la intención de ayudar a resolver un problema práctico significativo; la confusión y desinformación que pueden causar los textos de documentos públicos gubernamentales.

Al identificar y aplicar modelos que evalúen dicha complejidad léxica, se podrán desarrollar documentos sucesores simplificados, los cuales pueden reducir la necesidad de que los

ciudadanos se trasladen a las oficinas de las instituciones gubernamentales para que un empleado de dicha entidad les explique el comunicado o notificación ahorrando tiempo y dinero, como sugieren investigaciones recientes que indican que los modelos de lenguaje grande (LLMs) muestran un gran potencial en la simplificación de textos, aunque su potencial completo aún no ha sido aprovechado (Kew et al., 2023).

Este proyecto de investigación sirve de apoyo para los usuarios de la tercera edad, quienes enfrentan problemas para comprender textos gubernamentales, debido a elementos como la reducción de sus habilidades cognitivas y la comprensión del lenguaje técnico empleado en estos textos. Al simplificar la complejidad léxica de estos documentos, se les brinda un instrumento que les facilita el acceso a la información de forma más nítida y directa, fomentando de esta manera su independencia en la toma de decisiones vinculadas a los servicios del gobierno.

Este estudio no solo contribuirá a la práctica, sino que también sentará pautas en el campo del procesamiento del lenguaje natural (NLP). Al comparar el rendimiento de GPT-4 y sus sucesores, se generarán nuevos conocimientos sobre la eficacia de estas tecnologías en la simplificación de textos complejos. Como se señala en un estudio reciente, "gpt-4o muestra una mejora continua, de forma similar a las tendencias observadas en inglés, aunque las puntuaciones son más moderadas. Casi todos los modelos demuestran mejora cuando se acude a indicaciones más complejas" (Kelious et al., 2024), por lo que, puede influir en futuras investigaciones y aplicaciones en el área.

La solución que se derive de esta investigación tendrá múltiples utilidades. No solo beneficiará a los ciudadanos al facilitar la comprensión de documentos públicos, sino que también puede ser de gran utilidad para las instituciones gubernamentales, que podrán mejorar la claridad de su comunicación con los usuarios y, por ende, la eficiencia de sus procesos administrativos.

Limitaciones del estudio

Aunque el presente estudio es relevante, hay que tener en cuenta ciertas limitaciones que pueden afectar a su alcance y validez. A continuación, se mencionan las siguientes limitantes del estudio:

- *Acceso limitado a datos:* La disponibilidad de documentos públicos puede estar restringida por la falta de acceso a ciertas bases de datos o la confidencialidad de algunos documentos gubernamentales. Esto puede limitar la cantidad y diversidad de los textos utilizados para entrenar y evaluar los modelos.

- *Calidad de los datos:* La variabilidad en la calidad y formato de los documentos públicos puede requerir un esfuerzo mayor de preprocesamiento para estandarizar los datos, lo que puede generar sesgos o errores.
- *Capacidad de procesamiento:* Los modelos basados en la arquitectura Transformer, especialmente GPT-4, requieren recursos computacionales significativos para su entrenamiento y evaluación. Por lo que es importante tener un equipo de escritorio PC o laptop.
- *Tiempo de entrenamiento:* El tiempo necesario para entrenar y ajustar los modelos puede ser considerable, el tiempo de procesamiento de un modelo con una técnica es de 7 a 8 horas.
- *Comparación con otros modelos:* La comparación con otros modelos Transformer puede verse afectada por diferencias en la arquitectura, tamaño del modelo y metodologías de entrenamiento.
- *Costos:* El uso de las API de OpenAI es calculado por token, lo que significa que el saldo que se agrega a la API Key para la ejecución de los modelos resulta muy costoso.

CAPÍTULO II

MARCO TEÓRICO

Antecedentes del estudio

Con la ayuda de las TIC, las personas pueden acceder a una gran variedad de información en ámbitos tan diversos como la educación, la información, los asuntos sociales, entre otros. Según la autora (Belloch, 2011), numerosas personas no tienen acceso a esta información, puesto que algunas se topan con grandes barreras vinculadas a la lectura, tales como enunciados extensos, palabras poco comunes o estructuras lingüísticas complicadas que les dificultan entender el contenido de los textos.

Para el autor (Baquedano, 2006), los conceptos de legibilidad 'legibility' y lecturabilidad 'readability' se han utilizado como sinónimos en español, ambos refiriéndose a la facilidad o dificultad que presenta un texto para ser leído y comprendido. En este sentido (Thorndike, 1921) afirma que "leer no es sólo recrear la forma sonora de las palabras, sino también comprender", lo que resalta la importancia de la comprensión en el proceso de lectura.

Por otra parte (Tinker, 1963) argumenta que la legibilidad se centra únicamente en las características tipográficas del texto. (Alliende, 2002), en su obra "La legibilidad de los textos" sostiene que, la legibilidad se refiere a las características de los textos que favorecen o dificultan una comunicación más o menos eficaz entre ellos y los lectores. Dentro de este contexto, se puede considerar que una palabra compleja es aquella que resulta difícil de comprender para un usuario con un nivel bajo de alfabetización.

Desde un punto de vista más general, la predicción de la complejidad léxica (PCL) consiste en asignar una puntuación de complejidad a los términos, convirtiendo la tarea en un problema de regresión, en lugar de una tarea de clasificación binaria (J. A. Ortiz-Zambrano et al., 2023).

En estos últimos años, ha habido un considerable volumen de investigación en el área de la predicción de la complejidad textual y el procesamiento del lenguaje natural. A continuación, se mencionan investigaciones significativas que tratan estos aspectos:

"VYTEDU-CW: Difficult Words as a Barrier in the Reading Comprehension of University Students"(Ortiz Zambrano et al., 2020)

Metodología: Este análisis mostró la creación de un corpus multimodal que fusiona vídeos educativos y sus transcripciones, anotándolos con un nivel de complejidad del texto.

Resultados clave: El corpus "VYTEDU" se presentó como una herramienta útil para analizar la complejidad de los textos en contextos educativos, facilitando un estudio comparativo del discurso oral y escrito.

"SINAI at SemEval-2021 Task 1: Complex word identification using Word-level features" (J. Ortiz-Zambrano & Montejo-Ráez, 2021).

Metodología: Se utilizan datos de SemEval-2020 Task 1 para identificar palabras complejas en inglés, extrayendo características como longitud, frecuencia y embeddings preentrenados. Un clasificador Random Forest realiza la predicción, evaluando el rendimiento con el F1-Score.

Resultados clave: El modelo alcanzó un rendimiento competitivo en la identificación de palabras complejas, logrando una puntuación F1 de 0,85 en el conjunto de datos de evaluación, destacándose por la importancia de las características de frecuencia y embeddings.

"Transformers for Lexical Complexity Prediction in Spanish Language"(J. Ortiz-Zambrano et al., 2022).

Metodología: El artículo propone un enfoque para predecir la complejidad léxica en español utilizando modelos transformadores como BERT, XLM-RoBERTa y RoBERTa-large-BNE, entrenados con el corpus CLexIS2. Se combinan características manuales (frecuencia, longitud, categorías POS) con embeddings de modelos preentrenados.

Resultados clave: El modelo BERT ajustado alcanzó el rendimiento más destacado, con un MAE de 0.1592 y una correlación de Pearson de 0.9883. XLM-RoBERTa y RoBERTa-large-BNE igualmente mejoraron después del ajuste, aunque BERT resultó ser superior.

"SINAI Participation in SimpleText Task 2 at CLEF 2023: GPT-3 in Lexical Complexity Prediction for General Audience"(J. Ortiz-Zambrano et al., 2023).

Metodología: Se aplicaron los sucesores davinci-002 y davinci-003 del Modelo GPT-3 para la clasificación de la complejidad de las palabras y se utilizó el enfoque de aprendizaje few-shot, donde se proporcionaron ejemplos limitados al modelo para ayudar en la clasificación.

Resultados clave: El mejor rendimiento fue del modelo davinci-003 con un MAE de 0.0882, en la predicción de la complejidad se observaron coincidencias y discrepancias entre las categorías asignadas por GPT-3 y las del corpus Complex. Por último, se

identificaron oportunidades para explorar nuevos modelos como Claude 2 y GPT-4 en la predicción de la complejidad léxica.

"SINAI Participation in SimpleText Task 2 at CLEF 2024: Zero-shot Prompting on GPT-4-Turbo for Lexical Complexity Prediction"(J. Ortiz-Zambrano et al., 2024a).

Metodología: En esta investigación se empleó el modelo GPT-4 Turbo centrándose en dos subtareas: Sub-task 2.1 para la identificación de términos y asignaciones de niveles de dificultad y Sub-task 2.2 para la generación de definiciones y explicaciones de términos considerados como difíciles.

Resultados clave: Los resultados indicaron que GPT-4 Turbo mostro un rendimiento notable en la evaluación de la complejidad léxica sin necesidad de entrenamiento adicional. Para los resultados de Sub-task 2.1 se logró una buena capacidad para identificar términos relevantes y su dificultad y para Sub-task 2.2 las definiciones y explicaciones generadas fueron efectivas en términos difíciles mejorando la comprensión de los textos científicos.

"Deep Encodings vs. Linguistic Features in Lexical Complexity Prediction"(J. A. Ortiz-Zambrano et al., 2024)

Metodología: Para esta investigación se combinan características lingüísticas con codificaciones de modelos de lenguaje profundos (BERT, XLM-RoBERTa) en datasets en inglés y español, en los cuales se aplicaron varios algoritmos de aprendizaje automático.

Resultados clave: El modelo en inglés logró un MAE de 0.0683, mejorando un 29.2%, en cambio para el modelo en español se alcanzó un MAE de 0.1323, con una mejora del 19.4%.

En estudios anteriores como crear corpus multimodales y usar modelos avanzados como GPT-3 y GPT-4, muestran que es importante integrar características léxicas y semánticas en la identificación de palabras complejas. Estos estudios no solo han brindado recursos valiosos para la evaluación de la complejidad léxica, sino que también han brindado nuevas oportunidades para la mejora de la comprensión lectora en contextos educativos y científicos.

Fundamentación teórica

Los modelos de lenguaje natural se basan en redes neuronales y algoritmos de aprendizaje profundo para analizar, entender y generar el lenguaje humano. Por lo general, son capaces de aprender patrones complejos en grandes conjuntos de datos de texto, lo que les permite producir texto coherente de alta calidad. (Borjabad Salud .2023).

Aunque estos modelos han demostrado ser útiles en muchos ámbitos, el lenguaje humano tiene una capacidad reflexiva que le permite desarrollar una función metalingüística de la lengua, es decir, de poder reflexionar sobre diferentes aspectos, incluido el propio lenguaje natural, lo que aún falta desarrollar en estos modelos.

Por lo tanto, los fundamentos de esta investigación se basan en el procesamiento del lenguaje natural implementado en modelos de GPT, los factores que inciden en la complejidad textual de los documentos de índole público, entre otros que se indican a continuación con más detalle.

Procesamiento del Lenguaje Natural y Modelos Transformer

El Procesamiento del Lenguaje Natural (PLN), también conocido en inglés como Natural Language Processing (NLP) busca habilitar la comunicación fluida entre humanos y máquinas, el NLP constituye la base sobre el cual se desarrollan las aplicaciones de Inteligencia Artificial. (Angamarca, 2023) En este campo, se emplean modelos de lenguaje y algoritmos de aprendizaje automático para entrenar programas de computación capaces de entender, procesar y producir lenguaje humano. Según (Ramón & Corredera, 2023) el aprendizaje automático "se refiere a la asociación arquitecturas-algoritmos capaces de aprender las características de los datos de un dominio y, por tanto, de distinguir, de predecir y, en su caso, de generar de acuerdo con lo aprendido".

Quizás uno de los hitos más importantes de esta revolución surgió con la publicación del artículo "Attention Is All You Need" por investigadores de Google en 2017 (Vaswani et al., 2017). En este artículo se introdujo un cambio notable en la forma en que las computadoras comprenden y procesan el lenguaje humano, presentando una nueva arquitectura llamada Transformer. Según (Herrera Arjonilla, 2024) "estos modelos han revolucionado el ámbito del PLN, logrando comprender y producir texto con una coherencia y precisión que se asemejan cada vez más a las habilidades humanas".

Los modelos transformadores (Transformers) se componen de una estructura de redes neuronales creada específicamente para

manejar datos secuenciales, tales como texto o audio. Se expusieron en el artículo "Attention is All You Need" (Vaswani et al., 2017) y han revolucionado radicalmente el ámbito del procesamiento del lenguaje natural (NLP) y otras áreas vinculadas a la inteligencia artificial.

Además, resultan especialmente útiles para actividades que demandan entender el contexto y las conexiones a largo plazo presentes en el texto. Esto es apropiado para examinar la complejidad de documentos, dado que facilita la identificación de las sutilezas lingüísticas y estructurales que pueden complicar la comprensión de un texto. Como han señalado investigadores recientes, "los Transformers son capaces de modelar dependencias a largo plazo y comprender el contexto global de un texto, lo que les permite interpretar ambigüedades lingüísticas, identificar estructuras narrativas, etc." (Lewis Tunstall, 2022). Además, "modelos como ChatGPT y GPT-4 han demostrado el poder de los transformers en la generación de lenguaje natural, mostrando capacidades avanzadas en la conversación y la comprensión contextual, lo que abre nuevas posibilidades en aplicaciones de IA" (Liu et al., 2023).

Complejidad textual en documentos públicos

Los documentos públicos son parte de la comunicación entre el gobierno y la ciudadanía, por lo que es de vital importancia la comprensión de estos, sin embargo, la mayoría de veces resultan excesivamente complejos de entender para personas con diferentes niveles de educación y conocimientos técnicos. La falta de claridad en estos textos puede generar barreras en el acceso a la información, siendo así que afecta la participación ciudadana y el ejercicio de derechos.

El avance en los modelos de procesamiento del lenguaje natural, especialmente los de transformación, ha facilitado una evaluación más eficaz de esta complejidad textual, el autor (Díaz, 2024) señala que la característica más destacada de estos modelos es el mecanismo de atención, que les permite concentrarse simultáneamente en diferentes partes de la secuencia de entrada. En este sentido, durante esta sección se examinarán los factores que influyen en la complejidad textual y los métodos utilizados para evaluarla, centrándonos en el uso de tecnologías de procesamiento del lenguaje natural.

Factores que influyen en la complejidad textual

Según los autores (Alexandra Ortiz-Zambrano & Montejó-Raéz, 2020) en el artículo "Barriers in Reading Comprehension of University Students: Analysis of the Complicated Words Annotated in the VYTEDU-CW Corpus", entre los factores que contribuyen a la complejidad textual de los documentos públicos destacan los siguientes:

1. Palabras técnicas o especializadas
2. Estructuras sintácticas complejas
3. Densidad de la información
4. Abstracción de conceptos
5. Falta de coherencia e integridad del texto.

Así mismo, en otro artículo de (J. A. Ortiz-Zambrano & Montejó-Raéz, 2020) titulado "Overview ALexS 2020: CEUR Workshop Proceedings", se señala que la complejidad textual también puede verse influenciada por factores culturales y el conocimiento previo que el lector puede tener, especialmente en el caso de documentos legales y administrativos.

Los autores (Campos Saavedra et al., 2014), mencionan en su trabajo de investigación que los factores que influyen en la complejidad textual, como el léxico, la sintaxis y las interacciones entre el texto y el lector, son:

1. Propiedades estructurales y habilidades del lector
2. Factores lingüísticos
3. Complejidad léxica
4. Complejidad sintáctica
5. Modelo de comprensión lectora

Impacto de la complejidad textual en la participación ciudadana

La facilidad o dificultad con la que un texto puede ser leído y comprendido se conoce como lecturabilidad. (Campos Daniela et al., 2013). Esta dificultad en muchos casos se debe a la complejidad textual, siendo así que en los documentos públicos ejerce de manera directa un impacto en el ejercicio de los derechos, así como en la participación ciudadana.

Cuando los documentos públicos son difíciles de entender, se generan barreras como la desigualdad de acceso a la información, esto debido a que algunas personas con más conocimiento técnico y especializado logran entender, pero en otros casos no, esto desemboca en que hay menos probabilidad en que la ciudadanía en general participe en convocatorias, votaciones, consultas, entre otras.

Un caso de el impacto de la falta de comprensión se vio reflejado en la consulta del 21 de abril de 2024, donde la ciudadanía ecuatoriana fue citada a las urnas para votar sobre 11 preguntas, sin embargo, no toda la ciudadanía sabía exactamente sobre que se trataba las preguntas y el contexto e impacto que tenía su voto sobre esta consulta. La complejidad del lenguaje utilizado y la falta de información clara y accesible generaron confusión y desinformación entre los votantes, por lo que el voto, en muchos casos, fue emitido de manera desinformada o con incertidumbre.

Métodos de evaluación de complejidad textual

Hasta el momento, se evaluaba la complejidad de los textos mediante fórmulas de legibilidad y análisis lingüísticos. No obstante, la evolución del procesamiento del lenguaje natural ha posibilitado la elaboración de métodos más avanzados. Los autores (Ortiz Zambrano & Varela Tapia, 2019) en su estudio "Reading comprehension in university texts: The metrics of lexical complexity in corpus analysis in Spanish", han propuesto el siguiente enfoque multidimensional para evaluar la complejidad de los textos académicos:

1. Análisis léxico-semántico
2. Evaluación de la estructura sintáctica
3. Medición de la coherencia discursiva
4. Evaluación de la carga cognitiva

De igual manera, en otro artículo de (J. A. Ortiz-Zambrano & Montejo-Raéz, 2021) titulado "CLexIS2: A New Corpus for Identification Research in Computing Studies", se presenta un sistema fundamentado en el aprendizaje automático que toma en cuenta el contexto para calcular la complejidad de las palabras. Esto es especialmente significativo para documentos públicos que pueden incluir terminología especializada. Por otra parte, en el artículo "The BEA 2024 Shared Task on the Multilingual Lexical Simplification Pipeline" de los autores (Shardlow et al., 2024), se detallan los procedimientos esenciales empleados para valorar la complejidad textual, que incluyen la predicción de la complejidad léxica, tácticas de simplificación y métricas de evaluación utilizadas:

1. Pipeline de Simplificación Léxica Multilingüe (MLSP)
2. Predicción de Complejidad Léxica (LCP)
3. Baselines y métricas
4. Estrategias de ingeniería de prompts
5. Enfoques basados en características

Aplicación de modelos Transformer en la evaluación de complejidad textual

Existen modelos fundamentados en la arquitectura transformers que muestran un enorme potencial para valorar la complejidad textual. El artículo "Transformers for Lexical Complexity Prediction in Spanish Language at SEPLN 2022" (J. Ortiz-Zambrano et al., 2022) expone cómo estos modelos pueden ajustarse para prever la complejidad léxica en español. Similarmente, la investigación dirigida por Sarah Alissa de la Universidad de Manchester, titulada "Text simplification through Transformer and BERT" (Alissa & Wald, 2023), revela que la combinación de los modelos Transformer y BERT para simplificar textos complejos resulta efectiva. Este análisis investiga cómo estos métodos de procesamiento del lenguaje natural pueden potenciar de manera notable la comprensión de los documentos, con consecuencias directas para incrementar la disponibilidad de los textos de carácter público y educativo. Por otra parte, los autores (Rahali & Akhloufi, 2023) en su investigación "End-to-End Transformer-Based Models in Textual-Based NLP", hablan de varias aplicaciones de los modelos transformes en tareas de LPN como la clasificación de textos, donde se utilizan para categorizar textos en función de su nivel de dificultad o legibilidad, empleando modelos como BERT y sus sucesores.

Aplicaciones de simplificación textual en documentos

La tecnología ha desempeñado un papel clave al automatizar el proceso de adaptación de textos, permitiendo manejar grandes volúmenes de información y garantizar la accesibilidad para todos. Esta automatización optimiza la eficiencia, uniformiza la calidad y asegura que cada texto adaptado sea un elemento fundamental para la inclusión.

A pesar de lo mencionado anteriormente, según el autor (Alejandro Ayuso Expósito, 2024) simplemente simplificar el léxico no es suficiente, se requiere de una reorganización meticulosa y profunda que además de hacer la información legible sea comprendida por el lector. En este sentido, en un trabajo publicado por (M. R. García et al., 2012) da a conocer dos herramientas: La primera desarrollada para la simplificación léxica, la cual proporciona sinónimos para una palabra, dado que, una de las maneras más habituales de simplificar un texto es reemplazar palabras complejas por términos más simples.

La segunda herramienta trata la simplificación sintáctica, esta consiste en separar frases largas que contienen la conjunción "Y"

en frases más cortas, siguiendo la recomendación de distintas asociaciones en expresar una única idea en una sola frase.

La simplificación del lenguaje administrativo en la mejora de la accesibilidad y comprensión de documentos públicos.

La simplificación de textos es un proceso que ayuda a reducir la complejidad tanto de la redacción como de la estructura de la oración, al tiempo que se mantiene su significado. (Lasecki et al., 2015). Este proceso es especialmente relevante en herramientas de procesamiento del lenguaje natural (NLP) por ejemplo en traducción automática o en recuperación de información, donde la simplificación actúa como un paso de preprocesamiento clave, ya que las oraciones lingüísticas complejas se han identificado como un obstáculo. (Niklaus, C., 2016).

Dentro del ámbito jurídico según (González-Delgado & Navarro-Colorado, 2024), el lenguaje administrativo es aquel con el que la ciudadanía tiene mayor relación. Se trata del lenguaje producido por los organismos públicos para la aplicación de las leyes y normas jurídicas. Sin embargo, las estructuras oscuras y la terminología compleja pueden suponer una amenaza para la comprensión de su significado, impidiendo que las personas puedan realizar los trámites administrativos.

En el artículo "The Simplification of the Language of Public Administration: The Case of Ombudsman Institutions" publicado por los autores mencionados anteriormente se da a conocer ciertas recomendaciones para la redacción de documentos públicos:

1. Mantener oraciones breves, entre 15 a 20 palabras.
2. Utilizar términos claros y sencillos.
3. Evitar usar palabras complicadas o técnicas y optar por un sinónimo más sencillo.
4. Tener cuidado al momento de usar iniciales y acrónimos.
5. Utilizar la voz activa en lugar de construcciones pasivas e impersonales.
6. Dirigirse directamente al receptor.
7. Considerar utilizar ilustraciones, tablas y listas para facilitar la comprensión del material complejo.

GPT-4: Características y aplicaciones

GPT-4 es un modelo de estilo transformador entrenado previamente para predecir el siguiente token en un documento, utilizando tanto datos disponibles públicamente (como datos de Internet) como datos con licencia de proveedores externos. (Katz et al., 2024).

Según el artículo "GPT-4: A Review on Advancements and Opportunities in Natural Language Processing" (Baktash & Dawodi, 2023), este modelo cuenta con las siguientes características:

- *Capacidad multilingüaje*: Tiene mejores capacidades multilingües que sus predecesores, lo que le permite comprender y generar texto en múltiples idiomas.
- *Mejor comprensión textual*: Cuenta con una mejor comprensión del contexto en el que se genera el texto, lo que le permite generar textos más precisos y relevantes.
- *Mejor capacidad de razonamiento*: Su capacidad de razonamiento mejorada le permite realizar tareas complejas de razonamiento lógico.
- En el mismo artículo anterior, se mencionan las potenciales aplicaciones que tiene este modelo de IA, entre ellas tenemos:
 - *Chabots*: GPT-4 se puede utilizar para desarrollar Chabots más inteligentes que puedan comprender y responder a las consultas humanas con mayor precisión.
 - *Traducción de idiomas*: Este modelo puede ser utilizado para desarrollar modelos de traducción de idiomas más precisos y eficientes que puedan traducir texto entre varios idiomas.
 - *Resumen de texto*: Es posible desarrollar modelos de resumen de texto que puedan resumir grandes volúmenes de texto de forma precisa y eficiente.

GPT-4 y su aplicación en el análisis de complejidad textual

Los sistemas de conversación impulsados por IA han demostrado ser cada vez más efectivos para comprender y generar respuestas similares a las humanas en las comunicaciones basadas en texto, "a la vanguardia de esta revolución de la IA se encuentra Chat GPT 4, uno de los modelos de IA más populares y sofisticados desarrollados por OpenAI" (Van Guido, 2024).

GPT-4 es el último lanzamiento de la serie Generative Pre-trained Transformer (GPT), desarrollada por OpenAI que utiliza algoritmos de aprendizaje profundo para generar respuestas similares a las humanas a entradas basadas en texto. "GPT-4 promete ser aún más potente y versátil que sus predecesores, con el potencial de revolucionar la forma en que interactuamos con la tecnología" (Dhiman Devansh, 2023).

Este modelo se basa en el aprendizaje profundo con el fin de examinar el lenguaje de entrada y obtener el significado esencial. Durante el análisis de la estructura sintáctica, la semántica y el contexto, se pueden comprender las diferencias en el lenguaje y

generar respuestas más precisas y significativas. Cuenta con la integración de grafos de conocimiento para acceder a una amplia base de información. Según el autor mencionado con anterioridad “La integración permite aprovechar el conocimiento acumulado por la humanidad a lo largo de los años, haciéndolo más inteligente e informativo”.

En el contexto de nuestra investigación sobre la predicción de la complejidad de textos en documentos públicos, GPT-4 ofrece varias ventajas potenciales:

1. *Comprensión contextual*: GPT-4 puede analizar el contexto de un texto, esto le permite comprender el significado de las palabras y frases dentro de un contexto más amplio.
2. *Evaluar el estilo de la escritura*: Al evaluar el estilo de escritura; formal, informal, técnico, etc., puede proporcionar información adicional sobre la dificultad del texto.
3. *Detectar ambigüedades*: Al comprender el contexto puede detectar y analizar las posibles ambigüedades del lenguaje.
4. *Mejora continua*: Con el uso constante, GPT-4 puede aprender de las interacciones y mejorar su capacidad para predecir a complejidad textual.

Métodos tradicionales de evaluación de complejidad textual

Es importante contextualizar el uso innovador de GPT-4 para esta tarea. En este contexto, es importante revisar los métodos tradicionales de evaluación de la complejidad de los textos. Antiguamente se han utilizado varios métodos y enfoques, entre ellos:

1. *Fórmulas de legibilidad*: Como la fórmula Flesch-Kincaid o el índice Coleman-Liau, que se basan en características superficiales del texto, como la longitud de las frases y el número de sílabas por palabra.
2. *Análisis léxico*: Frecuencia y familiaridad de las palabras utilizadas en el texto.
3. *Análisis sintáctico*: Evalúa la complejidad de las estructuras gramaticales del texto.
4. *Coherencia y cohesión*: Herramientas que analizan la estructura lógica de las ideas presentadas en el texto y sus interrelaciones.

Aunque útiles, estos métodos tradicionales no suelen captar los aspectos más sutiles de la complejidad del texto, como el

contexto cultural o los conocimientos previos necesarios para comprenderlo.

Comparativa de modelos basados en la arquitectura Transformer

Para justificar la elección del GPT-4 en nuestro estudio, hay que realizar una comparativa con otros modelos basados en una arquitectura de Transformers. A continuación, se presenta una comparación entre los modelos GPT-4, GPT-3.5 y GPT-3.

1. GPT-4 (J. García, 2023)

- *Ventajas:* Puede procesar hasta 32.768 tokens (24.000 palabras aprox.) en un solo contexto, además mejora en la precisión y coherencia de las respuestas superando a sus predecesores, también tiene un manejo de instrucciones y tareas multimodales (texto e imágenes).
- *Desventajas:* Todavía presenta problemas de memoria, lo que limita su capacidad para realizar tareas complejas que requieran memoria a largo plazo.

2. GPT-3.5 (Trujillo et al., 2024) (J. García, 2023)

- *Ventajas:* Rendimiento mejorado en comparación con GPT-3 con una mayor capacidad de comprensión y generación de texto, a su vez permite una ventana de contexto de hasta 4.096 tokens (3.000 palabras aprox.).
- *Desventajas:* Su ventana de contexto es aún más limitada que la de GPT-4, lo que puede restringir su capacidad para manejar tareas que necesiten de una gran cantidad de información.

3. GPT-3 (Trujillo et al., 2024)

- *Ventajas:* Fue un modelo innovador que demostró un gran avance en la generación de texto coherente y relevante, tiene la capacidad para realizar tareas de LPN con pocos ejemplos (few-shot learning).
- *Desventajas:* Puede generar contenido sesgado o inapropiado, también tiene dificultad para interpretar decisiones y procesos internos (modelo de caja negra), además de poseer limitaciones en el manejo de contextos muy largos debido al costo computacional.

A raíz de esta comparación, se seleccionó GPT-4o y GPT-4 Turbo como sucesores de GPT-4 para nuestro estudio por su excelente capacidad para producir y comprender texto en una amplia gama de contextos, lo que resulta crucial a la hora de analizar documentos públicos complejos que pueden variar mucho en cuanto a estilo y contenido.

Estas premisas teóricas constituyen la base de nuestra investigación sobre la predicción de la complejidad del texto en documentos públicos utilizando GPT-4o y GPT-4 Turbo sucesores de GPT-4, a su vez proporcionan el marco necesario para comprender la novedad y el potencial de este enfoque en comparación con los métodos anteriores.

GPT-4 y su potencial en el análisis de documentos públicos

GPT-4 según (Ramalho et al., 2024) ha superado significativamente a sus predecesores en términos de capacidad y rendimiento, mejoró considerablemente en áreas como el razonamiento lógico, la generación de texto creativo y la comprensión contextual profunda, por esta razón puede abordar tareas más complejas y ofrecer respuestas más precisas y coherentes, lo que amplía su aplicabilidad en diversos dominios. Por otro lado, un estudio realizado por (Mao et al., 2023) señala que, aunque GPT-4 ha mostrado dificultades en tareas de inferencia del lenguaje natural, como manejar datos nuevos o ambiguos, su capacidad para identificar relaciones entre afirmaciones y generar desambiguaciones lo posiciona como una herramienta prometedora para analizar documentos públicos con lenguaje complejo.

GPT-4-Turbo para la predicción de la complejidad léxica

GPT-4 Turbo es una variante optimizada del modelo GPT-4, capaz de dar respuestas más rápidas en análisis de texto, generación de contenido y tareas que requieren evaluación y transformación de texto, como en el caso de la predicción de complejidad léxica.

En un estudio, se evaluó la capacidad de modelos de lenguaje grandes (LLMs) como GPT-4 Turbo para medir y modificar la legibilidad de textos en inglés, utilizando el corpus CLEAR. Los resultados mostraron que GPT-4 Turbo superó tanto a GPT-4o mini como a medidas tradicionales de legibilidad, logrando correlaciones positivas con juicios humanos. Además, el modelo fue capaz de reescribir textos para hacerlos más fáciles o más difíciles de leer, manteniendo su longitud original y obteniendo validaciones positivas de jueces humanos, lo que confirma su capacidad para ajustar la legibilidad de manera precisa y efectiva. (Trott & Rivière, 2024)

En otro estudio "SINAI Participation in SimpleText Task 2 at CLEF 2024: Zero-shot Prompting on GPT-4-Turbo for Lexical Complexity Prediction" (J. Ortiz-Zambrano et al., 2024), se demostró que este modelo es altamente competente en tareas de procesamiento del lenguaje natural, específicamente en la simplificación léxica y la identificación de palabras complejas, además fue capaz de generar ilustraciones detalladas creando así definiciones, explicaciones y casos de uso basados en el contexto de la oración.

GPT-4o para la predicción de la complejidad léxica

El GPT-4o es otro sucesor de GPT-4, cuyo lanzamiento fue en mayo de 2024. Se trata de una evolución del GPT-4 que mejora la velocidad, la eficacia y el procesamiento multimodal, lo que permite una interpretación más eficaz de texto, imágenes, audio y vídeo. En el contexto de la complejidad léxica, este modelo presenta avances significativos en la comprensión del lenguaje natural, lo que le permite interpretar y generar texto con una mayor precisión y coherencia. Esto incluye una mejora en la capacidad para manejar la ambigüedad y los matices del lenguaje.

En el artículo de Kelious, Constant y Coeur (2025), se utilizó GPT-4o para predecir la complejidad léxica en un contexto multilingüe, empleando diversas estrategias de prompting, como el zero-shot, one-shot y chain-of-thought. Evaluaron el rendimiento del modelo en tareas como la identificación de palabras complejas y la simplificación léxica multilingüe. Los resultados mostraron que GPT-4o alcanzó buenos resultados en algunos idiomas como alemán y español, especialmente con el prompting tipo Instruct, aunque su desempeño fue limitado en otros contextos y con el tipo de prompting Advanced COT (Kelious et al., 2024).

Zero-Shot y Few-Shot Learning

Zero-Shot Learning es una técnica del aprendizaje automático que permite a los modelos generar tareas hasta ahora desconocidas sin necesidad de ejemplos asignados (Alizadeh et al., 2023). Esto se consigue aprovechando el conocimiento general adquirido durante el entrenamiento, lo que permite al modelo clasificar o predecir basándose en datos que no ha visto antes.

Por otra parte, Few-Shot Learning es un paradigma del aprendizaje automático donde un modelo es entrenado utilizando un conjunto mínimo de ejemplos asignados para adaptar el modelo a nuevas tareas (Alizadeh et al., 2023),

usualmente se utilizan ejemplos de 2 a 100, comúnmente se emplea cuando es costoso o complicado recopilar grandes cantidades de datos etiquetados.

Zero-Shot Learning en Modelos de Lenguaje: Enfoque en GPT-4 y sus Sucesores

“Zero-Shot Learning (ZSL) permite a un modelo clasificar correctamente ejemplos de clases que nunca ha visto durante el entrenamiento. Esto se logra mediante la utilización de información auxiliar, como descripciones textuales o atributos de las clases” (Fernández Álvarez, 2024).

Esta capacidad es especialmente relevante en modelos de lenguaje como GPT. Según (García José, 2024), “los modelos GPT pueden realizar muchas tareas de procesamiento de lenguaje natural simplemente alterando las indicaciones del texto de entrada, sin entrenamiento específico para tareas individuales (conocido como ‘zero-shot learning’)”.

Esta versatilidad permite que modelos como GPT-4 y sus variantes manejen tareas de comprensión y generación de texto. Además, disminuye la necesidad de grandes conjuntos de datos etiquetados, ya que el modelo tiene la capacidad de utilizar descripciones textuales para guiar su desempeño.

Few-Shot Learning en Modelos de Lenguaje: Enfoque en GPT-4 y sus Sucesores

Según (Brown et al., 2020) few-shot learning funciona proporcionando K ejemplos de contexto y terminación, seguido de un ejemplo final de contexto, en el que se espera que el modelo proporcione la terminación correspondiente. Normalmente, fijamos K entre 10 y 100, ya que es el número de ejemplos que se pueden dar al modelo antes de realizar la predicción final.

Generalmente se emplea para realizar predicciones precisas con una cantidad limitada de ejemplo, según menciona el autor (Fernández Álvarez, 2024) few-shot learning, “es considerado una clase de meta-aprendizaje, es decir, el modelo aprende a aprender. El aprendizaje few-shot es un método efectivo de aprendizaje y muestra gran potencial”.

En el contexto de GPT-4 y sus sucesores, el uso de few-shot learning mejora la precisión y eficiencia del modelo, según el estudio del autor (Ono et al., 2024) el modelo GPT-4, en particular

su variante GPT-4Vision, ha demostrado habilidad para optimizar su desempeño en tareas de clasificación de imágenes y diagnóstico médico a través del uso de few-shot learning. Lo que posibilita a que el modelo se adapte de manera rápida a nuevas tareas con solo unos cuantos ejemplos.

En este sentido, el uso de few-shot learning en los sucesores del modelo GPT-4 puede aprovechar las características lingüísticas de los datos, como la complejidad léxica y el contexto para realizar predicciones más precisas con un número limitado de ejemplos.

Ética y consideraciones en el uso de modelos generativos para documentos públicos

El uso de los modelos generativos en documentos públicos plantea importantes cuestiones éticas y de responsabilidad. En el artículo “¿Didáctica con ChatGPT? Una propuesta ética y pedagógica” (Hernández-Fernández, A.,2023) se menciona que la tecnología emergente debe ser supervisada por una persona encargada, capaz de establecer límites claros en su aplicación y de asumir la responsabilidad por posibles errores o consecuencias no deseadas, así mismo se mencionan seis principios básicos a tener en cuenta:

- Prudencia.
- Fiabilidad.
- Rendimiento de cuentas.
- Responsabilidad.
- Autonomía limitada.
- No prescindir del humano.

Cabe mencionar la confiabilidad de la información generada por estos modelos, ya que según señala (Reichwein, F. L.,2024) si un contenido es percibido confiable, pero resulta ser falso o engañoso, podría debilitar aún más la confianza y generar mayor incertidumbre respecto a la información. Por esta razón, se hace evidente desarrollar e implementar estrategias eficaces para la verificación y la transparencia de los contenidos de los modelos generativos, a fin de garantizar que el público pueda confiar en la información que encuentra.

En el artículo titulado “Beyond principlism: Practical strategies for ethical AI use in research practices” (Lin et al., 2024) se da a conocer objetivos específicos para el uso ético de la IA, que se detallan a continuación:

1. Comprender el entrenamiento y el resultado del modelo, incluidas las estrategias de mitigación de sesgos.
2. Respetar la privacidad, la confidencialidad y los derechos de autor.
3. Evitar el plagio y las violaciones de políticas.
4. Utilizar la inteligencia artificial de manera que aporte beneficios en contraste con otras opciones.
5. Utilizar la IA de forma transparente y reproducible.

Variables de investigación e hipótesis

Las variables de esta investigación se componen de una variable independiente y una variable dependiente, que se detallan a continuación:

1. Variable independiente: Modelo Generative Pre-trained Transformer GPT-4.
2. Variable dependiente: Evaluación de la complejidad léxica en los textos públicos.

Hipótesis: El performance del modelo Generative Pre-trained Transformer GPT-4 y sus sucesores, incrementará la precisión de la predicción de la complejidad léxica en los textos públicos en español.

Revisión sistemática

Una revisión sistemática involucra un resumen crítico y reproducible de los resultados de las publicaciones disponibles sobre un mismo tema o pregunta clínica concreta, esto se realiza con el fin de mejorar la escritura científica. Para esta investigación se siguió el siguiente proceso:

Planteamiento de la pregunta estructurada

Las preguntas centrales de esta investigación son: ¿Que tan precisos son los modelos Generative Pre-Trained Transformer GPT-4 y sus sucesores para evaluar la complejidad léxica en documentos públicos? y ¿Cuáles son las características lingüísticas a tener en cuenta para lograr alcanzar el objetivo número uno de la investigación?

Búsqueda en bases de datos

Para la búsqueda se utilizaron bases de datos como Science Direct, ACL Anthology, ACM digital library, enfocándose en artículos publicados desde el 2015 hasta el 2024.

Selección de artículos

En esta sección se presentan a detalle los criterios para la selección de los estudios, la estrategia de exploración y los métodos para la recolección y síntesis de la información.

Para la revisión se considera los siguientes criterios:

Criterios de inclusión:

- Filtro por publicación: Artículos del área de tecnología, IA, lenguaje y gramática, artes lingüísticas, artículos de investigación de la evaluación semántica, comprendidos entre los años 2015 al 2024.
- Filtro por palabras: Entre las palabras claves están: Features, lexical complexity, modelo GPT, simplificación léxica, modelos de IA y el operador booleano AND. Los idiomas por tomar en cuenta para esta revisión son artículos tanto en español como en inglés.
- Ejemplo:
- "GPT-4" AND "lexical complexity"
- ("GPT-4o" OR "GPT-4 Turbo") AND "performance evaluation"
- "lexical features" AND "text analysis" AND "NLP" AND "complexity"

Criterio de exclusión:

- Estudios o investigaciones cuyo texto completo no esté disponible, ya sea por restricciones de acceso o por falta de respuesta de los autores.
- Publicaciones en idiomas distintos al español e inglés, ya que no serán considerados relevantes para esta revisión.
- Estudios o investigaciones que no incluyan las palabras clave especificadas o que no se relacionen directamente con los objetivos de la revisión.

Extracción de datos

Se realizó un análisis minucioso de cada una de las bases de datos y buscadores empleados, se evidencia que la temática puede dividirse en algunos campos, perspectivas, enfoques epistemológicos, tendencias o líneas de investigación como son: La inteligencia artificial, lenguaje y gramática, ciencias computacionales, procesamiento del lenguaje natural y análisis de complejidad textual.

Meta-análisis

Dentro del campo del análisis de complejidad textual e inteligencia artificial se encontraron varios artículos relacionados

al análisis y cálculo de características lingüísticas importantes para abrir el tema de la evaluación de la complejidad léxica en modelos GPT. Los principales hallazgos fueron:

- Características lingüísticas como: WordNet min-depth, WordNet max-depth (wn maxdepth). (Mosquera, 2021)
- Redacción acerca de la complejidad lingüísticas en documentos. (Ochoa Sierra & Cueva Lobelle, 2020)
- Puntos importantes acerca de la predicción de complejidad léxica en modelos de machine learning. (North et al., 2023)
- Métodos para predecir automáticamente la complejidad léxica en un entorno multilingüe utilizando modelos avanzados de procesamiento del lenguaje natural. (Choi et al., 2024)

Preguntas científicas a contestarse

Las preguntas científicas que se deben contestar con este estudio son las siguientes:

¿Hay una diferencia notable entre los modelos GPT-4, GPT-4 Turbo y GPT-4o al aplicar las técnicas de zero-shot learning y few-shot learning?

¿Qué tan consistentes son los resultados de predicción entre los diferentes sucesores del modelo GPT4 (GPT-4 Turbo y GPT-4o)?

Los modelos más recientes, como GPT-4 Turbo y GPT-4o, ¿Presentan mejoras significativas en comparación con GPT-4 en cuanto a precisión y rapidez en la predicción de complejidad léxica?

¿Son los modelos GPT-4 y sus sucesores realmente efectivos en la predicción de la complejidad léxica en textos públicos gubernamentales?

¿Pueden los modelos GPT-4 y sus sucesores reducir la complejidad léxica en documentos públicos para facilitar su comprensión por parte de los ciudadanos?

¿La implementación de estos modelos en la redacción de textos públicos podrían mejorar la transparencia y accesibilidad a la información gubernamental?

¿Es factible implementar una herramienta basada en GPT para que las instituciones públicas simplifiquen automáticamente sus documentos antes de publicarlos?

Definiciones conceptuales

Durante esta investigación se emplean palabras y conceptos claves para la comprensión del tema de estudio que se aborda, entre los principales tenemos:

Complejidad textual: La complejidad textual es el nivel de dificultad que presenta un texto para ser leído y comprendido. Según (Ochoa Sierra & Cueva Lobelle, 2020) la complejidad a nivel léxico se relaciona con la cantidad y variedad de términos, se dice que un texto con una amplia variedad de lexemas es más complejo que uno donde las mismas pocas palabras se repiten una y otra vez.

GPT 4: Es un modelo multimodal capaz de procesar imágenes y textos de entrada y producir textos de salida, así también cuenta con una gran capacidad para comprender y generar texto en lenguaje natural, en particular en situaciones complejas. (OpenAI, 2023)

Documentos públicos: Un documento público es el formado en el ejercicio de una actividad pública, es decir son los autorizados por un notario o empleado público, así como también los emitidos por instituciones públicas. (Díaz Rojas, 2005).

GPT-4 Turbo: Modelo de lenguaje avanzado desarrollado por OpenAI que puede analizar imágenes y proporcionar respuestas textuales a preguntas sobre ellas. (Khakzad Shahandashti kimya et al., 2024).

GPT-4o: Representa una actualización del previo GPT-4 de OpenAI .GPT-4o es un modelo omnidireccional autorregresivo que acepta como entrada cualquier combinación de texto, audio, imagen y video, y genera cualquier combinación de salidas de texto, audio e imagen. (OpenAI et al., 2024)

Few-Shot Learning: Es una variante de one-shot learning en la que el modelo se entrena con un número reducido de ejemplos (usualmente entre 2 y 100) por clase, con el fin de realizar predicciones precisas con pocos ejemplos (Fernández Álvarez, 2024).

Zero-Shot Learning: Permite que un modelo identifique correctamente ejemplos de categorías que no ha encontrado durante su entrenamiento, utilizando información adicional como descripciones o atributos de dichas categorías (Fernández Álvarez Cotutores et al., 2024).

CAPÍTULO III

METODOLOGÍA DE LA INVESTIGACIÓN

La metodología de investigación empleada en este estudio sobre la predicción de la complejidad léxica de textos en documentos públicos utilizando GPT-4 y modelos sucesores, se ha diseñado para abordar de manera integral los objetivos planteados. Este capítulo detalla la modalidad de la investigación, el tipo de investigación, así como el diseño metodológico.

Modalidad de la investigación

La modalidad de esta investigación es de carácter aplicada, ya que tiene como objetivo utilizar conocimientos existentes en procesamiento de lenguaje natural y en el análisis de modelos de inteligencia artificial para resolver un problema práctico: Evaluar la complejidad léxica de documentos públicos y mejorar su accesibilidad para los ciudadanos. Esta modalidad según (Vargas Cordero, 2009) incluye cualquier esfuerzo sistemático y socializado por resolver problemas o intervenir situaciones, ya que se concibe como investigación aplicada tanto la innovación técnica, artesanal e industrial como la propiamente científica.

De manera que esta investigación, permite vincular la teoría con la práctica, generando un impacto directo en la comprensión y uso de textos administrativos emitidos por las instituciones públicas de Guayaquil.

Tipo de investigación

El tipo de investigación de este estudio es cuantitativa y descriptiva, con un enfoque cuasi experimental. Este estudio es cuantitativo, porque se busca medir y analizar de manera numérica la complejidad léxica de los textos y evaluar el desempeño de los modelos GPT-4 y sus sucesores a través de métricas específicas como precisión, fluidez y coherencia.

En cuanto a porque se dice que es descriptivo, es debido a que se pretende caracterizar y analizar los elementos léxicos y lingüísticos presentes en los documentos públicos, permitiendo identificar patrones y tendencias.

Un enfoque cuasi experimental se dice que son estrategias que, guiadas por unos objetivos, pretenden examinar las relaciones existentes entre una o más variables independientes y la variable dependiente o de respuesta (Bono Cabré, 2012). Sabiendo lo anterior, se elige este enfoque porque se realizarán pruebas controladas con un corpus específico de documentos, simulando escenarios reales para evaluar los modelos en condiciones controladas.

Diseño metodológico de la investigación

El diseño metodológico de la investigación es esencial para garantizar que el estudio sea coherente. En esta sección se describe detalladamente el proceso para llevar a cabo la investigación que nos permitirá responder las preguntas de investigación y alcanzar los objetivos planteados.

Enfoque de la Investigación

La metodología del presente estudio es de tipo cuantitativo y descriptivo, con el fin de analizar y medir en primera instancia la complejidad léxica de documentos y notificaciones de las entidades que se han mencionado anteriormente mediante el cálculo de características lingüísticas, para de esta manera poder evaluar y analizar el desempeño de los modelos de GPT4 y sus sucesores en la predicción de la complejidad léxica.

Esta metodología fue elegida debido a su capacidad para proporcionar resultados objetivos y medibles, ya que esta metodología tiene la virtud de plantear una serie de pasos que permiten estudiar un fenómeno de manera estandarizada, acotando en gran medida la interferencia de sesgos (Hueso González & Cascant i Sempere, 2012).

El enfoque cuantitativo facilita la extracción de métricas lingüísticas específicas, mientras que el carácter descriptivo permite interpretar y contextualizar los datos obtenidos. De esta manera, se garantiza un análisis riguroso que no solo evalúe el rendimiento técnico de los modelos GPT-4 y sus sucesores, sino que también permita identificar patrones y tendencias en la complejidad léxica de los textos públicos emitidos por las entidades seleccionadas.

Además, se empleará un enfoque cuasi experimental para llevar a cabo pruebas controladas que permitan evaluar el desempeño de los modelos GPT-4 y sus sucesores en condiciones similares. Este enfoque implica el uso de un grupo de documentos, en particular oraciones previamente seleccionadas como muestra representativa, donde se analizará la complejidad léxica mediante métricas específicas generadas por los modelos.

Población

Para esta investigación la unidad de estudio corresponde a cada registro individual dentro del corpus que se tiene, el cual está compuesto por notificaciones e instrucciones relacionadas con trámites legales, en general oraciones seleccionadas de los documentos de las entidades públicas.

La población para el estudio como se menciona anteriormente incluye notificaciones e instrucciones de trámites legales recopiladas de cinco instituciones públicas gubernamentales de la ciudad de Guayaquil, estas son:

- Corporación Nacional de Telecomunicación (CNT).
- Servicio de Rentas Internas (SRI).
- Municipio de Guayaquil (MING).
- Autoridad de Tránsito Municipal de Guayaquil (ATM).
- Consejo Nacional Electoral (CNE).

En total, la población está conformada por un corpus de 7,813 registros.

Cada registro tiene los siguientes campos:

- id: Identificador único para cada registro.
- corpus: Institución pública gubernamental (Fuente).
- sentence: Oración que contiene la palabra etiquetada como compleja.
- token: Palabra identificada como compleja por etiquetadores.
- complexity: Valor numérico que representa la complejidad asignada por los etiquetadores.

Tabla
Corpus GovAIEc

3

id	corpus	sentence	token	complexity
607 5	Municipio - Tramites - TEXTO 0060 TRAMITES EN LA BIBLIOTECA MUNICIPAL.txt	- TRAMITES EN LA BIBLIOTECA MUNICIPAL Si en el stock existe el libro pedido, se emite Comprobante de la Publicación para que el comprador proceda a ...	comprobant e	0,33333333 3
771 9	SRI - Tramites - TEXTO 0130 REQUERIMIENTOS Y JUSTIFICACIONES DEL PROCESO INCONSITENCIAS.t xt	- REQUERIMIENTO S Y JUSTIFICACIONE S DEL PROCESO INCONSITENCIA S Si el Servicio de Rentas Internas detectare inconsistencias en las declaraciones o en los anexos que presente el contribuyente, siempre que no generen ...	sustitutivo	0,66666666 7
273 4	CNE - Tramites - TEXTO 0091 REGLAMENTO PARA CONFORMACION DE ALIANZAS ELECTORALES.txt	- REGLAMENTO PARA CONFORMACIO N DE ALIANZAS ELECTORALES la Constitución de la República dispone en el artículo 112, que los partidos y movimientos ...	militantes	1

Nota: La tabla muestra un conjunto de registros extraídos de documentos legales de diversas instituciones públicas de Guayaquil, con el objetivo de identificar palabras complejas dentro de los textos relacionados con trámites legales. Estos datos se utilizarán para calcular características lingüísticas y entrenar el modelo.

Elaborado por: Scarlet Gutierrez e Ivette Zamora.

La escala de complejidad tiene los siguientes niveles:

- Moderately difficult: Rango de complejidad entre 0 y 0.3333. Las palabras u oraciones en este rango son algo complejas, pero se pueden entender en su contexto.
- Difficult: Desde 0.3334 a 0.6666. Las palabras u oraciones en este nivel son bastante complejas y pueden necesitar un mayor nivel de comprensión o conocimientos técnicos.
- Very difficult: Desde 0.6667 a 1. Las palabras u oraciones en esta categoría son muy complejas, lo que puede hacer que sean difíciles de entender.

Tabla 4

Escala de complejidad

Etiqueta	Rango
moderately difficult	(0, 0.3333)
difficult	(0.3334, 0.6666)
very difficult	(0.6667, 1)

Nota: Esta escala se utiliza para evaluar el nivel de complejidad de las palabras en el prompt, empleando las técnicas zero-shot learning y few_shot learning para los modelos GPT-4. Además, esta escala fue utilizada por los anotadores para asignar un valor a la palabra identificada como compleja (token), el cual se registró en la columna de complexity. **Elaborado por:** Scarlet Gutierrez e Ivette Zamora.

Características lingüísticas

Las características lingüísticas que se usaran para calcular la complejidad en los textos de los documentos públicos están dadas por 40 características en total.

Las primeras 15 características son mencionadas en el artículo científico "SINAI at SemEval-2021 Task 1: Complex word identification using Word-level features" (J. Ortiz-Zambrano & Montejo-Ráez, 2021), siendo estas:

1. **Frecuencia absoluta:** Se refiere a la frecuencia o cantidad de veces con la que se presenta un determinado fenómeno en una población. (Granados, 1994). En esta investigación se usará esta característica para calcular el número de veces que se repite el token (palabra compleja).
2. **Frecuencia relativa:** Para los autores (Lúcar Gonzalez & Fernando, 2019) la frecuencia relativa es "la probabilidad de obtener cierto dato, se obtiene calculando la razón entre la frecuencia absoluta de un dato con el total. Se puede expresar como fracción, decimal o porcentaje". En esta investigación, la frecuencia relativa se utilizará para calcular la proporción de veces que un token (palabra compleja) aparece en relación con el total de tokens que incluye el corpus, permitiendo evaluar su importancia o prevalencia dentro del texto estudiado.
3. **Longitud de la palabra:** Los autores del artículo en donde se encontró esta característica, se refieren al termino longitud de palabra, a el número de caracteres de la palabra objetivo (token). Como lo menciona (Armas et al., 2016) las palabras largas requieren más tiempo y esfuerzo para ser procesadas y pronunciadas, lo que aumenta la carga cognitiva. Al tener que recordar varias palabras largas, se experimenta una mayor demora o dificultad al intentar recuperarlas de la memoria. Por el contrario, las palabras más cortas suelen ser más fáciles de recordar y procesar, ya que requieren menos tiempo para ser almacenadas y recuperadas; y es por esta razón que se toma en cuenta esta característica ya que nos permitirá visualizar la relación entre la longitud de las palabras y su impacto en el rendimiento cognitivo, lo que afecta a la comprensión.
4. **Número de sílabas:** La sílaba es "una unidad de procesamiento en español". (Álvarez et al., 1998). Esta característica será aplicada al token del corpus, con el fin de analizar cómo la estructura silábica de las palabras puede influir en su procesamiento y comprensión. El mismo autor menciona que al contar el número de sílabas de cada palabra, se puede observar si las palabras con más sílabas requieren mayor tiempo de procesamiento o si afectan la facilidad con la que los lectores las reconocen y retienen.
5. **Posición del token:** Posición de la palabra objetivo en la oración. Según (Tsaprouni & Manouilidou, 2025) la posición del token en una oración, es decir, la ubicación de una palabra objetivo dentro de la estructura sintáctica de la oración se calcula principalmente para entender cómo la disposición de las palabras afecta su procesamiento cognitivo, su facilidad de comprensión y el impacto en la memoria de los lectores.

6. **Número de palabras en la oración:** Se calcula el número de palabras en la oración que se encuentra en el corpus, con el fin de ver cómo la longitud de la oración afecta la carga de la memoria de trabajo durante la lectura. El autor (De et al., 2009) menciona que, al incrementar el número de palabras, se incrementa la cantidad de información que debe ser retenida y procesada activamente, lo que puede influir en la capacidad de los individuos para realizar tareas cognitivas complejas, como la comprensión lectora y el razonamiento.
7. **Parte del discurso:** Parte del discurso o part of speech según (Petrov et al., 2011) se refiere a un término utilizado en lingüística para describir las categorías en las que se agrupan las palabras de acuerdo con su función dentro de una oración. Este autor también menciona que una oración con un gran número de sustantivos o verbos puede ser más informativa, mientras que una con muchas preposiciones y pronombres podría ser más compleja en términos de estructura, pero menos densa en información.
8. **Frecuencia relativa de la palabra antes de la palabra objetivo (token):** Se calcula la frecuencia relativa de la palabra que se encuentra posicionada en la oración antes del token, ya que la frecuencia de las palabras previas a una palabra objetivo ofrece contexto lingüístico, facilitando la comprensión de su significado y las interacciones entre las palabras en una oración. (Sims-Williams, 2022)
9. **Frecuencia relativa de la palabra después de la palabra objetivo (token):** Asimismo como se menciona en la característica anterior se calcula la frecuencia relativa de la palabra después de la palabra objetivo, ya que al igual que la anterior característica proporciona información clave sobre el contexto en el que se encuentra la palabra objetivo.
10. **Longitud de la palabra anterior:** El número de caracteres en la palabra antes del token. Esta característica se tomó en cuenta, por que como lo menciona (Jalbert et al., 2011) puede afectar la facilidad con que se accede a la palabra objetivo en la memoria.
11. **Longitud de la palabra que sigue:** El número de caracteres en la palabra después del token. Esta característica proporciona un contexto adicional para interpretar la palabra objetivo. (Barton et al., 2014)
12. **Medida de diversidad léxica textual:** La diversidad léxica de la palabra objetivo en la oración utilizando la métrica planteada por (McCarthy & Jarvis, 2010), estos indican que esta medida se refiere a un indicador que se utiliza para evaluar la variedad de palabras o vocabulario presente en un texto. Además, mencionan

que una mayor diversidad léxica tiende a ser más complejo el texto.

13. **Número de sinónimos:** Un sinónimo según lo menciona (Phoocharoensil, 2010) se refiere a una relación de similitud o igualdad de significado entre dos o más palabras. De acuerdo con lo mencionado por (White, 2014) contar los sinónimos en un texto permite analizar cómo el autor utiliza variaciones léxicas para ajustar el tono, la formalidad y el estilo, influenciando así la complejidad percibida del texto.
14. **Número de hipónimos:** “Un hipónimo es aquella palabra que posee todos los rasgos semánticos de otra más general” (Ortega-Mendoza et al., 2007). En el contexto de esta investigación calcular los hipónimos en la oración presente en el corpus nos ayuda en el análisis de cómo los textos en documentos públicos emplean términos específicos que podrían ser complejos para ciertos lectores.
15. **Número de hiperónimos:** “Son palabras que en su significado incluyen a otra u otras (por ejemplo, fruta respecto de banana o naranja)” (Hael, 2016). Se toma en cuenta esta característica ya que podría ser una herramienta para medir si los textos usan términos muy generales, que pueden ser adecuados para el público general, pero tal vez carezcan de precisión para ciertos contextos.

Adicionalmente se encontraron las 15 características anteriormente mencionadas junto con 8 nuevas características lingüísticas en el artículo “LegalEc: A New Corpus for Complex Word Identification Research in Law Studies in Ecuatorian Spanish” (J. A. Ortiz-Zambrano et al., 2023), las cuales son:

16. **Número de sustantivos singular o plural:** Como señala (Bosque, 2016) un sustantivo “es la parte de la gramática con la que se designan seres materiales o inmateriales de muy diversa naturaleza”, a su vez este mismo señala que un sustantivo singular se refiere a un solo ser o material, mientras que un sustantivo plural hace referencia a muchos seres materiales o inmateriales. Cabe recalcar que el cálculo de esta característica permitirá verificar si la oración presentada en el corpus contiene sustantivos técnicos o abstractos que puedan perjudicar la comprensión.
17. **Número de verbos auxiliares:** El autor (Pottier, 1961) menciona que “auxiliar es todo verbo que es ‘incidente’ de otro verbo en un mismo sintagma verbal “. Esta característica será calculada en el campo de sentence (oración) del corpus con el fin de indicar el nivel de complejidad sintáctica del texto, ya que según menciona el autor un uso excesivo de auxiliares podría generar estructuras poco claras, especialmente en textos administrativos o técnicos,

lo que puede dificultar la comprensión para lectores no especializados.

18. **Número de adverbios:** Se refiere a el número de adverbios presentes en la oración, en este caso en la oración presente en el corpus. El adverbio según (Torner Castells & Dra Paz Battaner Dra Louise McNally, 2005) se refiere a una palabra que expresa circunstancias en el que se presenta un evento, ya que ofrece información acerca del tiempo, lugar en donde se desarrolla la acción, entre otros. Para esta investigación calcular el número de adverbios puede ser útil para evaluar cómo los textos públicos utilizan este tipo de palabras para lograr claridad, precisión o persuasión.
19. **Número de símbolos:** "El símbolo se define como aquel signo cuyo significado representa cierto signo de otro plano o de otra lengua." (Lotman et al., 2002). Al calcular esta característica en la oración se podrá verificar si existen símbolos y determinar su impacto en la interpretación del texto, evaluando cómo contribuyen a la transmisión de significados.
20. **Número de expresiones numéricas:** "Una "expresión numérica" como una frase que expresa una cantidad que puede venir acompañada de un modificador numérico como, por ejemplo, casi un cuarto" (Bautista et al., 2012). Se calcula el número de expresiones numéricas que existen en la oración, estas expresiones pueden resultar complejos para personas que no están acostumbradas a interpretar cantidades o datos estadísticos.
21. **Número de verbos:** Según lo menciona (De Miguel, 2004) el verbo "sería en principio una palabra dotada de la posibilidad de seleccionar semántica y formalmente uno o dos complementos (sintagmas que complementen o completen su significado)". Evaluar su presencia en la oración del corpus ayuda a determinar si las ideas están bien desarrolladas o si presentan redundancias.
22. **Número de sustantivos:** Esta característica se calcula para el campo *sentence* presente en el corpus que corresponde a la oración.
23. **Número de pronombres:** Los pronombres hacen referencia a palabras que reemplazan al sustantivo en una oración y te evitan tener que mencionarlo directamente. (Dapena, 1986). Se calcula con el fin de analizar la cohesión, referencia textual y claridad del mensaje al sustituir o evitar la repetición de sustantivos.

Finalmente, se hallaron 17 características lingüísticas adicionales. Las primeras 8 características se encontraron en el artículo científico "SemEval-2021 Task 1: Exploring Sentence and Word Features

for Lexical Complexity Prediction” (Mosquera, 2021), entre estas están:

24. **Número de morfemas:** El autor (González Calvo, 1990) define al morfema como “la unidad lingüística recurrente mínima que tiene un significado y no puede ser analizada en unidades mínimas significativas recurrentes menores”. Esta característica nos permitirá evaluar si existen morfemas en la estructuración de las palabras de la oración, de esta manera se puede identificar si los textos están escritos con un lenguaje comprensible.
25. **Longitud del lema:** En el artículo donde se haya esta característica, se indica que esta se refiere a la cantidad de caracteres en la forma básica o canónica de una palabra. Además (Strauss et al., 2007) señala que es importante en diversas tareas de procesamiento de lenguaje natural (PLN), ya que permite estandarizar las palabras a su forma base para facilitar el análisis.
26. **Is stopword (Es una palabra vacía):** Según el artículo donde se menciona esta característica, is stopword, hace referencia a que es verdadero si la palabra de destino es una palabra vacía. Una palabra vacía como lo menciona (Kaur, 2018) o también conocidas como palabras de ruido, son las palabras que contienen poca información que normalmente no se necesita. Además, este autor menciona que esto se utiliza para mejorar la eficiencia de los modelos por la razón de que no aportan suficiente información para el análisis semántico y, por lo tanto, son comúnmente eliminadas durante la tokenización y otros procesos de preprocesamiento de texto.
27. **Número de sentidos de una palabra:** En el artículo se menciona que es el número de sentidos de la palabra de destino extraída de WordNet. El autor (Fellbaum, 2010) indica que “WordNet es una gran base de datos léxica electrónica para el inglés”, también señala la utilidad de esta característica en tareas de procesamiento de lenguaje natural (PLN) como la desambiguación semántica, que busca determinar el significado correcto de una palabra en un contexto determinado.
28. **Índice de legibilidad de Flesch:** Es una métrica que evalúa la facilidad de lectura de un texto en función de la longitud promedio de las oraciones y la cantidad promedio de sílabas por palabra. (Barrio Cantalejo & Simón Lorda, 2003). Esto nos permitira evaluar la facilidad de lectura del texto en documentos públicos.
29. **Índice de Gunning-Fog:** Es una métrica de legibilidad que estima la cantidad de años de educación formal que una persona necesita para comprender un texto en su primera lectura. (Larbi et al., 2023). Esta característica nos permitirá inferir que grado de

educación formal deben tener las personas para comprender el texto en el documento público.

30. **Índice de SMOG:** Es una métrica de legibilidad que estima los años de educación necesarios para comprender un texto de manera cómoda. (Zhou et al., 2017). Esta métrica nos ayudara a saber los años de educación que son necesarios para comprender el texto de los documentos públicos.
31. **Índice RIX:** Es una métrica de legibilidad que se calcula en función del número de oraciones y la longitud de las palabras en el texto. (Alfonso-Robaina et al., 2011). Esta nos ayudara a obtener una idea general de la complejidad de un texto en cuanto a su facilidad de lectura.

Otras 5 características que se suman a esta lista se encontraron en el artículo científico "UWB at SemEval-2016 Task 11: Exploring Features for Complex Word Identification" (Konkol, 2016).

32. **n-gramas de caracteres:** Se refiere a un conjunto de n elementos consecutivos en un documento de texto, que puede incluir palabras, números, símbolos y puntuación. (Stamatatos, 2009). Según la autora citada anteriormente esta característica es útil para la detección de plagio, la clasificación de texto o el reconocimiento de patrones.
33. **WordNet synset size:** El autor del artículo en donde se señala esta característica, indica que "el tamaño de los conjuntos de sinónimos de WordNet que contienen la palabra actual. Los conjuntos de sinónimos grandes pueden ser muy ambiguos y, por lo tanto, más difíciles para un hablante no nativo."
34. **WordNet number of synset:** El mismo autor señala que esta característica se refiere a "la cantidad de conjuntos de sinónimos que contienen la palabra actual. Las palabras con muchos significados pueden resultar más difíciles."
35. **Language model sentence probability. (Probabilidad de oración del modelo de lenguaje):** El autor se refiere a esta característica como "la probabilidad general de una oración que puede indicar la complejidad de la oración. Una oración compleja puede afectar la dificultad de la palabra actual."
36. **Average n-gram frequency. (Frecuencia promedio de n-gram):** El autor indica que "si la oración contiene una mayor cantidad de palabras complejas, entonces incluso una palabra común puede volverse compleja."

Las siguientes dos características fueron seleccionadas del artículo "The Whole is Greater than the Sum of its Parts: Towards the Effectiveness of Voting Ensemble Classifiers for Complex Word Identification" (Wani et al., 2018):

37. **Degree of Polyseny o número de sentidos de la palabra objetivo en WordNet:** El autor indica que esta característica "se hace operativo contando el número de conjuntos de sinónimos de la palabra de destino en WordNet. Las palabras con tamaños mayores de conjuntos de sinónimos de WordNet tienen varios sentidos y se descubrió que son más confusas."
38. **Número de vocales:** Una vocal es aquella que se produce sin requerir una gran apertura de la cavidad bucal, por lo que la lengua se coloca en una posición más elevada en la boca, sin llegar a generar una fricción en el sonido. (Llisterri, 1989). Esta característica será aplicada al token y nos permitirá obtener información sobre las características fonológicas del token y su posible impacto en la fluidez o ritmo del discurso.

Otra característica fue hallada en el artículo "LCP-RIT at SemEval-2021 Task 1: Exploring Linguistic Features for Lexical Complexity Prediction" (Desai et al., 2021):

39. **Word complexity lexicon:** Se refiere a un conjunto de palabras asociadas a un valor o puntuación que indica la complejidad de cada palabra en términos de su facilidad de comprensión o uso. (Maddela & Xu, 2018) . Esto nos ayudara a ver la complejidad del token en términos matemáticos.

Como ultima característica se encuentra Phrase lenght in terms of words and characters o longitud de una frase en términos de palabras y caracteres, se refiere a la cantidad de palabras y caracteres que contiene una frase, hallada en el artículo "Phase evolution of Peregrine-like breathers in optics and hydrodynamics" (Xu et al., 2019)

Algunas de las características lingüísticas mencionadas se refieren a métricas específicas que tienen fórmulas definidas, y las otras características son de naturaleza más descriptiva o cuantificable sin una fórmula matemática exacta. Estas características que involucran fórmulas son principalmente las métricas de legibilidad y aquellas relacionadas con la frecuencia de aparición de palabras en un texto.

De las 17 características lingüísticas adicionales identificadas, 4 de ellas tienen fórmulas específicas que permiten calcular sus valores, estas son:

- Índice de legibilidad de Flesch (Flesch Reading Ease),
- Índice RIX (RIX Score),
- Índice de Gunning-Fog (Gunning-Fog Index) y
- Índice de SMOG (SMOG Score).

A continuación, se presenta una tabla con las fórmulas de estas características lingüísticas que corresponden a las métricas de legibilidad de un texto:

Tabla**5***Fórmulas de las métricas de legibilidad*

Nombre	Fórmula
	Ecuación 1 Índice de legibilidad de Flesch
Flesch Reading Ease	$FRE = 206.835 - 1.015 \left(\frac{\text{total words}}{\text{total sentences}} \right) - 84.6 \left(\frac{\text{total syllables}}{\text{total words}} \right)$ Tomado de: (Kher et al., 2017)
	Ecuación 2 Índice RIX
RIX Score	$RIX = \frac{\text{number of long words}}{\text{number of sentence}}$ Tomado de: (Anderson, 1983)
	Ecuación 3 Índice de Gunning-Fog
Gunning-Fog Index	$GFI = 0.4 \left[\left(\frac{\text{total words}}{\text{total sentences}} \right) + 100 \left(\frac{\text{complex word}}{\text{total words}} \right) \right]$ Tomado de: (Kher et al., 2017)
	Ecuación 4 Índice de SMOG
SMOG Score	$SMOG = 1.043 \sqrt{\text{number of complex words} * \frac{30}{\text{total sentences}}} + 3.1291$ Tomado de: (Kher et al., 2017)

Nota: Las fórmulas presentadas en esta tabla son utilizadas para evaluar la legibilidad de un texto, proporcionando métricas que permiten analizar su complejidad y accesibilidad. **Tomado de:** (Anderson, 1983), (Kher et al., 2017).

Elaborado por: Scarlet Gutierrez e Ivette Zamora.

Los resultados de estas fórmulas tienen un gran significado que permite la interpretación de la complejidad del texto. Para cada una de las características lingüísticas correspondientes a las métricas de legibilidad de un texto se presenta una tabla con el rango de estos valores y su interpretación. Los niveles educativos mencionados en todas las tablas son adaptaciones de un artículo de investigación en EE.UU. (Kher et al., 2017) y han sido ajustados para reflejar los niveles de educación en Ecuador.

A continuación, se presenta una tabla que muestra la clasificación de dificultad de lectura según la métrica Flesch Reading Ease. Esta métrica evalúa la legibilidad de un texto y proporciona una puntuación que indica el nivel educativo aproximado necesario para comprenderlo.

Tabla

6

Clasificación de dificultad de lectura - Flesch Reading Ease

Puntuación	Interpretación (Nivel Educativo Aproximado en Ecuador)	Clasificación de Dificultad de Lectura (USDHHS)
0-29	Universitario (Postgrado)	Difícil
30-49	Universitario (Licenciatura)	Difícil
50-59	Bachillerato	Difícil
60-69	Secundaria (8°-9° grado)	Moderado
70-79	Secundaria (7° grado)	Moderado
80-89	Primaria (6° grado)	Fácil
90-100	Primaria (5° grado)	Fácil
Valor negativo	El texto puede ser incomprensible para la mayoría de los lectores.	Muy difícil

Nota: Esta tabla es una adaptación del sistema educativo de Ecuador, tomando los niveles educativos de EE.UU de la tabla original. **Tomado de:** (Kher et al., 2017)

Elaborado por: Scarlet Gutierrez e Ivette Zamora.

A continuación, se presenta la tabla de la clasificación de dificultad de lectura según la métrica RIX Score, la cual utiliza la misma escala que la métrica Flesch Reading Ease.

Tabla**7***Clasificación de dificultad de lectura - RIX Score*

Puntuación	Interpretación (Nivel Educativo Aproximado en Ecuador)	Clasificación de Dificultad de Lectura (USDHHS)
0-29	Universitario (Postgrado)	Difícil
30-49	Universitario (Licenciatura)	Difícil
50-59	Bachillerato	Difícil
60-69	Secundaria (8°-9° grado)	Moderado
70-79	Secundaria (7° grado)	Moderado
80-89	Primaria (6° grado)	Fácil
90-100	Primaria (5° grado)	Fácil
Valor negativo	El texto puede ser incomprensible para la mayoría de los lectores.	Muy difícil

Nota: Esta tabla es una adaptación del sistema educativo de Ecuador, tomando los niveles educativos de EE.UU de la tabla original. **Tomado de:** (Kher et al., 2017)

Elaborado por: Scarlet Gutierrez e Ivette Zamora.

La siguiente tabla corresponde a la métrica de legibilidad Gunning-Fog Index, esta métrica también evalúa la legibilidad de un texto y proporciona una puntuación que indica el nivel educativo aproximado necesario para comprenderlo.

Tabla**8***Clasificación de dificultad de lectura - Gunning-Fog Index*

Puntuación	Interpretación (Nivel Educativo Aproximado en Ecuador)	Clasificación de Dificultad de Lectura (USDHHS)
0-6	Primaria (hasta 6° grado)	Fácil
6-10	Secundaria (7°-10° grado)	Moderado
10+	Bachillerato o Educación Superior	Difícil
Valor negativo	El texto puede ser incomprensible para la mayoría de los lectores.	Muy difícil

Nota: Esta tabla es una adaptación del sistema educativo de Ecuador, tomando los niveles educativos de EE.UU de la tabla original. **Tomado de:** (Kher et al., 2017)

Elaborado por: Scarlet Gutierrez e Ivette Zamora.

Por último, tenemos la tabla de SMOG Score que al igual que la métrica de legibilidad Gunning-Fog Index comparten la misma escala, esta métrica también evalúa la legibilidad de un texto y

proporciona una puntuación que indica el nivel educativo aproximado necesario para comprenderlo.

Tabla

9

Clasificación de dificultad de lectura - SMOG Score

Puntuación	Interpretación (Nivel Educativo Aproximado en Ecuador)	Clasificación de Dificultad de Lectura (USDHHS)
0-6	Primaria (hasta 6° grado)	Fácil
6-10	Secundaria (7°-10° grado)	Moderado
10+	Bachillerato o Educación Superior	Difícil
Valor negativo	El texto puede ser incomprensible para la mayoría de los lectores.	Muy difícil

Nota: Esta tabla es una adaptación del sistema educativo de Ecuador, tomando los niveles educativos de EE.UU de la tabla original. **Tomado de:** (Kher et al., 2017)

Elaborado por: Scarlet Gutierrez e Ivette Zamora.

Según (Farrús et al., 2005) la normalización min-max es la técnica más simple: los valores mínimos y máximos de los scores se desplazan a los valores 0 y 1, respectivamente, y todos los scores se transforman en el rango [0,1], de manera que la distribución original se mantiene (excepto para el factor de escala). En este sentido, luego de calcular todas las características lingüísticas al corpus "GovAIEc", a los resultados obtenidos se aplicó la normalización utilizando el método Min-Max, que escala los valores de las características para que se encuentren en un rango de 0 a 1. Este enfoque se implementó para asegurar que los datos sean comparables y para evitar que los resultados se vean afectados por picos o valores extremos.

Figura

1

Función de Normalización

```
# Normalización de las columnas
def normalize_column(column):
    """Normaliza una columna de pandas."""
    return (column - column.min()) / (column.max() - column.min())
    if (column.max() - column.min()) != 0 else 0
```

Nota: El fragmento del código muestra la función `normalize_column`, que normaliza la columna de la característica lingüística, escalando sus valores entre 0 y 1. Esto se realiza al restar el valor mínimo de la columna y dividirlo entre el rango de la columna (máximo - mínimo).

Elaborado por: Scarlet Gutierrez e Ivette Zamora.

Grupos experimentales

En este contexto, los “grupos experimentales” son los diferentes modelos de predicción que se comparan en cuanto a su rendimiento. No son grupos de sujetos, sino modelos entrenados y evaluados sobre el mismo corpus. Los grupos experimentales son:

- GPT-4: Este modelo basado en la arquitectura Transformer se entrenó para predecir la complejidad textual en documentos públicos.
- GPT-4 Turbo: Versión del modelo GPT-4 utilizada para predecir la complejidad de los textos.
- GPT-4o: Otra versión del modelo GPT-4 que se compara con los otros modelos para determinar su efectividad en la tarea de predecir la complejidad textual.

Tratamiento (Modificación)

En el tratamiento de los datos para la identificación de palabras complejas del corpus GovAIEc, se utilizó un enfoque basado en el modelo de lenguaje GPT-4 (uso de zero-shot learning y few-shot learning). Este enfoque fue empleado para predecir la complejidad textual de las palabras en función de su contexto dentro de las oraciones extraídas de los documentos legales. El tratamiento de los datos consistió en los siguientes pasos:

1. *Lectura y preprocesamiento de datos:* Los datos fueron leídos desde el corpus GovAIEc.xlsx que contenía las oraciones y las palabras identificadas como complejas en los documentos legales de las instituciones públicas. En particular, las columnas id, sentence, token y complexity fueron los campos importantes para calcular la predicción del modelo y mostrar el resultado.

2. *Generación de predicciones de complejidad:* Para la clasificación de las palabras en función de su complejidad, se emplearon los modelos basados en la arquitectura Transformer, específicamente GPT-4, GPT-4 Turbo y GPT-4o. Estos modelos se aplicaron para identificar el nivel de complejidad de las palabras dentro de las oraciones de los documentos legales.

Cada modelo utiliza un enfoque de Zero-Shot Learning, lo que significa que no necesita ejemplos etiquetados previamente para realizar las predicciones. El modelo clasifica cada palabra identificada como compleja en una de las tres categorías de complejidad.

Además, también se empleó para estos modelos la técnica de Few-Shot Learning, ya que les permite aprender con pocos ejemplos proporcionados en el prompt, en específico se proporcionaron un total de 20 ejemplos, con el objetivo de mejorar la precisión de la predicción. Los sucesores del modelo GPT-4 clasifican cada palabra identificada como compleja en una de las tres categorías de complejidad.

3. *Formato del prompt de predicción:* El modelo recibe como entrada un prompt específico que establece el contexto de la tarea.

A continuación, se muestra el formato del prompt utilizado para realizar las predicciones de complejidad mediante Zero-Shot Learning:

Figura

2

Prompt Zero-Shot Learning

```

samples > zero_shot_prompt.txt
prompt = (
  "I am reading fragments of texts from public institution sources, and some words are difficult to understand. "
  "I want to automatically identify complex words and determine their level of difficulty. "
  "Please classify the word provided into one of the following categories based on the sentence context: "
  "\"moderately difficult\" (value: from 0 to 0.3333), \"difficult\" (value: from 0.3334 to 0.6666), or \"very difficult\" (value: "
  "from 0.6667 to 1). "
  "After reading the fragment: '@oracion', I find that the word '@aEvaluar' is classified as one of these categories. "
  "Return only the category name with no extra explanation or content. "
  "The output should only be one of the following options: 'moderately difficult', 'difficult', or 'very difficult'."
)

```

Nota: El código muestra el prompt utilizado para clasificar palabras complejas en tres categorías de dificultad basándose en el contexto de la oración. Esta clasificación se realiza mediante Zero-Shot Learning. **Elaborado por:** Scarlet Gutierrez e Ivette Zamora.

A continuación, se muestra una parte del formato prompt utilizado para realizar las predicciones de complejidad mediante Few-Shot Learning:

Figura

3

Prompt Few-Shot Learning

```

samples > few_shot_prompt.txt
prompt = (
  "I am analyzing fragments of texts from public institution sources in Spanish. Some words in these texts are difficult to "
  "understand, and I want to automatically identify complex words and classify their difficulty levels. "
  "The difficulty levels are: "
  "- \"Moderately difficult\" "
  "- \"Difficult\" "
  "- \"Very difficult\" "
  "The criteria for classification include factors such as word length, frequency of use in everyday language, and the technical or "
  "formal nature of the word. "
  "Here is an example text fragment: "
  "\"TRÁMITES EN LA BIBLIOTECA MUNICIPAL. Si en el stock existe el libro pedido, se emite comprobante de la Publicación para que el "
  "comprador proceda a realizar el pago en las Cajas Recaudadoras (Palacio Municipal, Malecón 1600 y Clemente Ballén, esq. Bloque "
  "Noreste 1, planta baja).\" "
  "From this text, I found that the word \"comprobante\" is classified as \"moderately difficult.\" "
  "Your task is to: "
  "1. Identify complex words in the text. "
  "2. Classify each complex word as \"moderately difficult,\" \"difficult,\" or \"very difficult.\" "
  "3. Return the words and their classifications in a structured format, such as a list or table. "
  "\n\n##\n\n"
)

```

Nota: El fragmento de código muestra parte del prompt utilizado para clasificar palabras complejas en tres categorías de dificultad basándose en el contexto de la oración. Esta clasificación se

realiza mediante Few-Shot Learning. **Elaborado por:** Scarlet Gutierrez e Ivette Zamora.

El código toma la oración en la que se encuentra la palabra compleja y el token de esa palabra, y el modelo devuelve la categoría de complejidad correspondiente.

4. *Evaluación y resultados:* Una vez que el modelo realiza las predicciones de complejidad, los resultados se almacenan en un Excel con el nombre resultado_gpt4turbo_version_zero_shot_Test.xlsx, resultado_gpt4o_version_zero_shot_Test.xlsx, resultado_gpt4_version_zero_shot_Test.xlsx, resultado_gpt4turbo_version_few_shot_Test.xlsx, resultado_gpt4o_version_few_shot_Test.xlsx y resultado_gpt4_version_few_shot_Test.xlsx para su posterior análisis.

5. *Resumen final:* Luego de tener todos los resultados, se genera un Excel a manera de resumen final con todos los resultados de las métricas para cada modelo y técnica empleada, el archivo se guarda con el nombre resumen_metricas.xlsx.

Test Estadísticos

Las fórmulas estadísticas aplicadas en esta investigación corresponden a métricas de evaluación utilizadas para medir la precisión y calidad de las predicciones realizadas por el modelo GPT-4 y sus sucesores GPT-4 Turbo y GPT-4o. A continuación, se mencionan cada una de ellas:

- MAE (Mean Absolute Error): Puede utilizarse si los valores atípicos representan partes corruptas de los datos. (Chicco et al., 2021)

Ecuación 5 Mean Absolute Error

$$MAE = \frac{1}{m} \sum_{i=1}^m |X_i - Y_i|^2$$

Tomado de: (Chicco et al., 2021)

- MSE (Mean Squared Error): Se emplea para evaluar la exactitud de un modelo de predicción, es una evaluación de la diferencia entre los valores detectados y los valores estimados por el modelo. También puede utilizarse cuando es necesario detectar valores atípicos, debido a la norma L2, el MSE otorga un mayor peso a estos puntos, si el modelo produce una única predicción muy mala, la parte cuadrática de la función incrementa el error. (Chicco et al., 2021)

Ecuación 6 Mean Squared Error

$$MSE = \frac{1}{m} \sum_{i=1}^m (X_i - Y_i)^2$$

Tomado de: (Chicco et al., 2021)

- RMSE (Root Mean Squared Error): MSE y RMSE mantienen una relación monótonica, es decir, por medio de la raíz cuadrada. Una ordenación de los modelos de regresión basada en el MSE será idéntica a una ordenación de los modelos basada en el RMSE. (Chicco et al., 2021)

Ecuación 7 Root Mean Squared Error

$$RMSE = \sqrt{MSE}$$

Tomado de: (Botchkarev, 2018)

- R^2 (R-squared): "El coeficiente de determinación puede interpretarse como la proporción de la varianza de la variable dependiente que puede predecirse a partir de las variables independientes." (Chicco et al., 2021, p. 5)

Ecuación 8 R-squared

$$R^2 = 1 - \frac{\sum_{i=1}^m (X_i - \bar{Y}_i)^2}{\sum_{i=1}^m (\bar{Y} - \bar{Y}_i)^2}$$

Tomado de: (Chicco et al., 2021)

- Pearson Correlation Coefficient: Usualmente se utiliza la correlación de Pearson para medir la correlación entre dos variables cuando la variable objetivo a medir es de naturaleza constante. (Gallego, 2024)

Ecuación 9 Pearson Correlation Coefficient

$$\rho = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

Tomado de: (Shong & Bs, 2008)

- Spearman Correlation Coefficient: "El coeficiente de correlación Spearman se puede usar como alternativa al coeficiente de correlación Pearson ya que puede ser menos sensible a datos atípicos al usar una clasificación de rangos de los datos originales, evaluando una relación monótonica pero no necesariamente lineal como sucede con el coeficiente de Pearson." (Gallego, 2024, p. 24)

Ecuación 10 Spearman Correlation Coefficient

$$r = \frac{\sum_{i=1}^m ((X_i - \bar{X})(Y_i - \bar{Y}))}{\sqrt{\sum_{i=1}^m (X_i - \bar{X})^2 \sum_{i=1}^m (Y_i - \bar{Y})^2}}$$

Tomado de: (Shong & Bs, 2008)

Los coeficientes de correlación; Pearson y Spearman son herramientas estadísticas importantes para evaluar la relación entre dos variables. La interpretación de estos coeficientes es importante para entender la fuerza y la dirección de la relación observada. A continuación, se presenta una tabla que resume la magnitud absoluta de estos coeficientes de correlación y su correspondiente interpretación.

Tabla**10**

Interpretación de Coeficientes de Correlación

Magnitud absoluta del coeficiente de correlación observado	Interpretación
0.00-0.10	Correlación negligible
0.10-0.39	Correlación débil
0.40-0.69	Correlación moderada
0.70-0.89	Correlación fuerte
0.90-1.00	Correlación muy fuerte

Nota: Esta tabla presenta la interpretación convencional de los coeficientes de correlación, tanto para el coeficiente de Pearson como para el coeficiente de Spearman. Los rangos de magnitud absoluta del coeficiente de correlación se utilizan para evaluar la fuerza de la relación entre dos variables. **Tomado de:** (Schober & Schwarte, 2018)

Elaborado por: Scarlet Gutierrez e Ivette Zamora.

Estas métricas son aplicadas para evaluar el rendimiento del modelo GPT-4 y sus de manera cuantitativa, al calcular estas métricas podemos determinar que tan bien el modelo está percibiendo los valores de complejidad en comparación con los valores reales.

En este sentido, las métricas se calculan utilizando las columnas complexity (valores reales) y complejidad GPT-variante de GPT-4 (valores predichos) del corpus GovAIEc. Lo cual, es importante para entender la efectividad del modelo y realizar ajustes o mejoras en caso de ser necesario, estas métricas nos permiten comparar el modelo GPT-4 y sus sucesores para ver cuál es el más eficiente en la tarea de la identificación de palabras complejas.

Para evaluar la diferencia relativa entre los modelos en términos de porcentaje se utilizará la fórmula de diferencia porcentual, la cual es mencionada por (Yuen et al., 2024) y se expresa de la siguiente manera:

Ecuación 11 Diferencia porcentual

$$\% = \left(\frac{x_1 - x_2}{x_2} \right) \times 100$$

En donde:

- x_1 representa el valor final.
- x_2 representa el valor inicial.

Figura

4

Cálculo de Métricas de Evaluación para Predicción de Complejidad

```
# calcular métricas
true = resultado.loc[:, "complexity"]
predicted = resultado.loc[:, "Complejidad GPT-4-turbo"]

metrics = {
    "MAE": round(mean_absolute_error(true, predicted), 6),
    "MSE": round(mean_squared_error(true, predicted), 6),
    "RMSE": round(mean_squared_error(true, predicted, squared=False), 6),
    "R2": round(r2_score(true, predicted), 6),
    "Pearson": round(true.corr(predicted, method='pearson'), 6),
    "Spearman": round(true.corr(predicted, method='spearman'), 6)
}

for m in metrics:
    resultado[m] = metrics[m]
    print(f"{m}: {metrics[m]}")
print("\n")
```

Nota: El fragmento de código muestra el cálculo de las métricas de evaluación para comparar la predicción de complejidad generada por el modelo, en este caso GPT-4-turbo con la complejidad real. Se calculan las métricas estadísticas MAE, MSE, RMSE, R², y las correlaciones de Pearson y Spearman para comparar la precisión del modelo. **Elaborado por:** Scarlet Gutierrez e Ivette Zamora.

Beneficiarios directos e indirectos del proyecto

Los principales beneficiarios directos serán los ciudadanos que interactúan con documentos públicos emitidos por las instituciones seleccionadas para el estudio. Este grupo incluye a personas que enfrentan dificultades para comprender el lenguaje técnico o burocrático utilizado en notificaciones e instrucciones relacionadas con trámites legales. La implementación de los modelos GPT-4 y sus sucesores para evaluar estos textos permitirá a los ciudadanos acceder a información más clara y

comprensible, facilitando así la realización de trámites de manera más eficiente y autónoma.

Así como también, serán beneficiados directamente las instituciones públicas gubernamentales seleccionadas para el estudio, al contar con herramientas que analicen la claridad de sus documentos. Esto les permitirá identificar áreas de mejora en su escritura, optimizando la interacción con la ciudadanía y reduciendo el número de consultas o errores derivados de la mala interpretación de sus textos.

Otros de los beneficiarios directos son los investigadores de esta rama, ya que los datos e información que se obtenga de este estudio puede servir como base para futuras investigaciones relacionadas con procesamiento de lenguaje natural (NLP), accesibilidad lingüística y simplificación textual.

Finalmente, como beneficiarios indirectos tenemos a la ciudadanía en general del país, aunque el proyecto solo se realice con información e instituciones a nivel de la ciudad de Guayaquil su implementación podría extenderse a nivel nacional. Por otra parte, las instituciones académicas podrían utilizar los resultados y enfoque del proyecto como material de enseñanza en cursos sobre inteligencia artificial, procesamiento de lenguaje natural y accesibilidad textual.

Entregables del proyecto

El presente proyecto se centra en el análisis de la complejidad léxica de textos públicos utilizando modelos avanzados de procesamiento del lenguaje natural, en específico el modelo GPT-4 y sus sucesores. Para evaluar debidamente la complejidad léxica, se fijaron objetivos específicos. En este sentido, se definieron varios entregables para documentar cada etapa del proyecto con el fin de garantizar la transparencia y replicabilidad de los resultados. A continuación, se presenta el siguiente listado de los entregables del proyecto:

- Documento de Tesis
- Corpus GOVAIEC Preprocesado
- Código fuente de la aplicación de los Modelos
- Resultados finales de Evaluación

Propuesta

Esta tesis se enfoca en la evaluación de la complejidad léxica de textos públicos utilizando modelos avanzados de procesamiento de lenguaje natural GPT-4 y sus modelos sucesores GPT-4 Turbo y GPT-4o, así como también el descubrimiento de características lingüísticas que permiten verificar la complejidad léxica del corpus titulado como "GovAIEc". Los modelos de IA

mencionados anteriormente fueron entrenados con el corpus a fin de que puedan calcular la complejidad del token y así poder comparar si concuerdan con el campo de complejidad asignado en el corpus, además de comparar el rendimiento de los modelos con las dos técnicas de aprendizaje; zero-shot learning y few-shot learning.

Criterios de validación de la propuesta

Para la validación de la propuesta presentada en esta investigación, se recurrió a la evaluación por parte de un panel de expertos en las áreas de informática. Este proceso tuvo como objetivo garantizar que la investigación cumpla con los requisitos definidos y que su funcionamiento sea eficiente y conforme a los estándares establecidos.

Tabla

11*Expertos*

N°	Nombre	Título Profesional	Años de experiencia
1	Kenia Lisbeth Nieves Pincay	Ingeniero de Software	4
2	Emilio Jesús Beltrán Valle	Ingeniero en sistemas de información	4
3	Henry Miguel Ruiz Reyes	Ingeniero de Software	4
4	Rubén Hernán Vera Párraga	Ingeniero de Software	4

Nota: Esta tabla presenta a los expertos que, con su experiencia y criterio, se encargarán de evaluar el proyecto.

Elaborado por: Scarlet Gutierrez e Ivette Zamora.

Para este juicio de expertos se planteó que se evalúen los siguientes criterios de selección, con el fin de garantizar que la propuesta no solo sea técnicamente sólida, sino también pertinente y aplicable en el contexto de los documentos públicos.

Tabla

12*Criterios de evaluación*

Indicador	Criterios
Eficiencia del Modelo	Capacidad del modelo para identificar y clasificar la complejidad léxica en textos públicos.
Pertinencia y Aplicabilidad	Adecuación del modelo a los textos públicos específicos.

Metodología de Evaluación	Solidez y claridad de la metodología utilizada para evaluar el desempeño del modelo.
Impacto Social	Potencial para mejorar la comprensión de documentos públicos entre ciudadanos.

Nota: Esta tabla presenta los criterios de evaluación que serán medidos por los expertos mencionados anteriormente.

Elaborado por: Scarlet Gutierrez e Ivette Zamora.

El resultado de los juicios de expertos se presenta en la siguiente tabla, en la cual se detalla el puntaje obtenido por cada criterio de evaluación de la propuesta.

Tabla

13

Juicio de expertos

Indicador	Nombre del experto	Puntuación	Observaciones
Eficiencia del Modelo	Kenia Lisbeth Nieves Pincay	95	Ninguna
	Emilio Jesús Beltrán Valle	90	Ninguna
	Henry Miguel Ruiz Reyes	100	Ninguna
	Rubén Hernán Vera Párraga	100	Ninguna
Pertinencia y Aplicabilidad	Kenia Lisbeth Nieves Pincay	95	Ninguna
	Emilio Jesús Beltrán Valle	95	Ninguna
	Henry Miguel Ruiz Reyes	100	Ninguna
	Rubén Hernán Vera Párraga	100	Ninguna
Metodología de Evaluación	Kenia Lisbeth Nieves Pincay	95	Ninguna
	Emilio Jesús Beltrán Valle	100	Ninguna
	Henry Miguel Ruiz Reyes	100	Ninguna
	Rubén Hernán Vera Párraga	100	Ninguna
Impacto Social	Kenia Lisbeth Nieves Pincay	100	Ninguna
	Emilio Jesús Beltrán Valle	95	Ninguna
	Henry Miguel Ruiz Reyes	100	Ninguna
	Rubén Hernán Vera Párraga	100	Ninguna

Nota: Esta tabla presenta las puntuaciones otorgadas por cada experto para los indicadores establecidos. **Elaborada por:** Scarlet Gutierrez e Ivette Zamora.

Resultados

Características lingüísticas

Los resultados del cálculo de las 40 características se presentan a continuación mediante los campos de media, mediana, valor mínimo (Min) y valor máximo (Max):

Tabla

14

Características lingüísticas

Característica	Media	Mediana	Min	Max
Frecuencia absoluta	0,05	0,02	0	1
Frecuencia relativa	0,05	0,02	0	1
Longitud de la palabra	0,35	0,35	0	1
Número de sílabas	0,33	0,33	0	1
Posición del token	0,16	0,14	0	1
Número de palabras en la oración	0,24	0,2	0	1
Número de sinónimos.	0,01	0	0	1
Número de hipónimos.	0	0	0	1
Número de hiperónimos.	0,01	0	0	1
Parte del discurso	0,11	0,12	0	0,88
Frecuencia relativa de la palabra antes de la palabra objetivo	0,23	0,06	0	1
Frecuencia relativa de la palabra después de la palabra objetivo	0,25	0,14	0	1
Longitud de la palabra anterior	0,22	0,14	0	1
Longitud de la palabra que sigue	0,16	0,12	0	1
mtld_diversity	0,21	0,19	0	1
Número de pronombres.	0,24	0,17	0	1
Número de verbos auxiliares.	0,19	0,17	0	1
Número de verbos.	0,18	0,15	0	1
adp	0,23	0,2	0	0,82
Número de sustantivos.	0,29	0,27	0	0,93
nn	0	0	0	0
sym	0,01	0	0	1
num	0,06	0	0	1
Número de morfemas	0,17	0,2	0	1
Longitud del lema	0,33	0,35	0	1
Is stopword	0,46	0,42	0	1

Número de sentidos de una palabra	0,13	0,06	0	1
Índice de legibilidad de Flesch	0,64	0,66	0	1
Índice de Gunning-Fog	0,28	0,25	0	1
Índice de SMOG	0,36	0,33	0	1
Índice RIX	0,27	0,23	0	1
n-gramas de caracteres	0,24	0,2	0	1
WordNet synset size.	0,06	0	0	1
WordNet number of synset.	0,07	0	0	1
Language model sentence probability	0	0	0	1
Average n-gram frequency.	0,05	0	0	1
Degree of Polyseny	0,07	0	0	1
Número de vocales.	0,47	0,44	0	1
Word complexity lexicon	0,37	0,37	0	1
longitud de una frase en términos de palabras y caracteres	0,24	0,2	0	1

Nota: Esta tabla presenta de forma resumida los resultados del cálculo de 40 características lingüísticas extraídas del corpus, con el fin de analizar diversos aspectos relacionados con las palabras, oraciones y su contexto. Para cada característica, se han calculado las siguientes estadísticas descriptivas: **media**, **mediana**, **valor mínimo (Min)** y **valor máximo (Max)**. Los valores en cada columna corresponden a un resumen estadístico de los datos de las características lingüísticas para el corpus analizado, lo cual permite observar las variaciones y tendencias de las mismas a través de los diferentes textos presentes en el conjunto de datos.

Elaborado por: Scarlet Gutierrez e Ivette Zamora.

Aplicación del modelo GPT-4 y sus sucesores GPT-4 Turbo y GPT-4o

A cada modelo se le aplicó las técnicas de zero-shot learning y few-shot learning. Como se mencionó anteriormente, para cada ejecución se guardó el corpus procesado según la técnica empleada. A continuación, se presentan los resultados de cada modelo junto con la técnica utilizada. Los modelos seleccionaron palabras que definieron como complejas a partir de las diferentes oraciones presentadas en el corpus.

A continuación, se presentarán unas tablas con un pequeño extracto de los resultados obtenidos. Se han tomado las palabras clasificadas como difíciles de la columna "up to 5 difficult terms

retrieved by GPT-4-turbo” por el modelo GPT-4 y sus sucesores utilizando las técnicas zero-shot learning y few-shot learning para cada una de las instituciones públicas gubernamentales. Es decir, el modelo evalúa la dificultad de comprensión del texto y marca 5 palabras difíciles de entender para una persona promedio, estas son las palabras que se presentaron en las tablas. Primero presentaremos las palabras complejas del modelo GPT-4, en el cual se utilizó la técnica de zero-shot learning a continuación se presentan las palabras:

Tabla**15***Palabras complejas GPT-4 / Zero-Shot Learning*

Institución	Palabras complejas
MIMG	Referendum, Constitucional, Directrices, Contingencia, Proyección
CNE	Resolución, Aprobó, Desagregados, Ejecución, Estableciéndose
SRI	Requerimientos, Justificaciones, Superintendencia, Representante, Contribuyentes
CNT	Cesionario, Calificación, Notificada, Proceda, Reglamentario
ATM	Jurídicas, Normativo, Autónomos, Descentralizados, Prestación

Nota: La tabla presenta algunas palabras que el modelo GPT-4 aplicando la técnica de zero-shot learning clasificó como complejas. **Elaborado por:** Scarlet Gutierrez e Ivette Zamora.

Para el modelo GPT-4 Turbo, utilizando la técnica de zero-shot learning, se presentan las siguientes palabras:

Tabla**16***Palabras complejas GPT-4 Turbo / Zero-Shot Learning*

Institución	Palabras complejas
MIMG	Jurídicas, Municipal, Publicaciones, Alcaldía, Donación
CNE	Desafiliación, Nulidad, Afiliación, Registrada, Requerirlo
SRI	Requerimientos, Justificaciones, Inconsistencias, Contribuyente, Administración
CNT	Institución, Irrevocable, Verídica, Ostentar, Cesionaria
ATM	Proyectarán, Peatonales, Multimodales, Integren, Atracciones

Nota: La tabla presenta algunas palabras que el modelo GPT-4 Turbo aplicando la técnica de zero-shot learning clasificó como complejas. **Elaborado por:** Scarlet Gutierrez e Ivette Zamora.

En cuanto al modelo GPT-4o aplicando zero-shot learning el modelo identificó las palabras como difíciles:

Tabla**17***Palabras complejas GPT-4o / Zero-Shot Learning*

Institución	Palabras complejas
MIMG	Naturales, Jurídicas, Donación, Publicaciones, Solicitando
CNE	Desafiliación, Electrónico, secretaria, Asignado, Seguimiento
SRI	Requerimientos, Justificaciones, Notificación, Administración, Institucional

CNT	Superintendencia, Socioeconómica, Cumplimiento, Información, Adicionales
ATM	Proyectaran, Recorridos, Peatonales, Conecten, Atracciones

Nota: La tabla presenta algunas palabras que el modelo GPT-4o aplicando la técnica de zero-shot learning clasificó como complejas. **Elaborado por:** Scarlet Gutierrez e Ivette Zamora.

En el caso del modelo GPT-4 con la técnica de few-shot learning, se muestran las siguientes palabras:

Tabla

18

Palabras complejas GPT-4 / Few-Shot Learning

Institución	Palabras complejas
MIMG	Ocurre, Registrada, Convenio, Débito, Vigente
CNE	Calificará, Organizaciones, Solicitudes, Contratación, Publicidad
SRI	Requerimientos, Justificaciones, Proceso, Contribuyente, Tributaria
CNT	Institución, Irrevocable, Verídica, Ostentar, Cesionaria
ATM	Tacómetro, Sonómetro, Frenómetro, Luxómetro, Regloscopio

Nota: La tabla presenta algunas palabras que el modelo GPT-4 aplicando la técnica de few-shot learning clasificó como complejas. **Elaborado por:** Scarlet Gutierrez e Ivette Zamora.

Para GPT-4 Turbo con la técnica few-shot learning, se muestran las siguientes palabras:

Tabla**19***Palabras complejas GPT-4 Turbo / Few-Shot Learning*

Institución	Palabras complejas
MIMG	Comisario, Municipal, Compromiso, Determinado, Subsanan
CNE	Ratifico, Autorización, Expiración, Notificar, Cancelar
SRI	Requerimientos, Justificaciones, Resolución, Carácter, Información
CNT	Telecomunicaciones, Tranquilidad, Autorización, Determinadas, Propaguen
ATM	Estupefacientes, Psicotrópicas, Terapéutico, Rehabilitación, Infracciones

Nota: La tabla presenta algunas palabras que el modelo GPT-4 Turbo aplicando la técnica de few-shot learning clasificó como complejas. **Elaborado por:** Scarlet Gutierrez e Ivette Zamora.

Por último, para el modelo GPT-4o aplicando few-shot learning el modelo identificó las siguientes palabras como difíciles:

Tabla**20***Palabras complejas GPT-4o / Few-Shot Learning*

Institución	Palabras complejas
MIMG	Préstamo, Consultado, Instalaciones, Reteniendo, Identificación
CNE	Desafiliación, Renuncia, Escaneado, Nulidad, Seguimiento
SRI	Conformación, Organizaciones, Discriminatorias, Funcionamiento, Garantizara
CNT	Prestación, Establecidos, Reparaciones, Situaciones, Fortuito
ATM	Inminente, Seguridad, Ocupantes, Obligación, Comprobar

Nota: La tabla presenta algunas palabras que el modelo GPT-4o aplicando la técnica de few-shot learning clasificó como complejas. **Elaborado por:** Scarlet Gutierrez e Ivette Zamora.

A continuación, se presentan los resultados obtenidos a manera de resumen de la aplicación del modelo GPT-4 y sus sucesores GPT-4 Turbo, GPT-4o en la identificación de palabras complejas. Para evaluar la precisión de las predicciones de complejidad generadas por estos modelos, se utilizaron las métricas estadísticas ya descritas; MAE, MSE, RMSE, R^2 , y las correlaciones de Pearson y Spearman para determinar la efectividad de los modelos en la clasificación de la complejidad léxica. Lo que nos permite una comparación cuantitativa entre los distintos sucesores y su rendimiento en la tarea de simplificación del lenguaje. Los resultados se muestran en la siguiente tabla:

Tabla **21**

Resultado de los sucesores

0	MAE	MSE	RMSE	R2	Pears on	Spear man	Coincide ncia
GPT- 4_few_shot	0,246 41	0,088 819	0,298 025	- 0,593 562	0,074 48	0,0569 53	37,59
GPT- 4Turbo_few _shot	0,254 051	0,091 523	0,302 527	- 0,642 07	0,315 69	0,3093 87	51,67
GPT- 4o_few_sho t	0,259 385	0,096 301	0,310 325	- 0,727 808	0,316 272	0,3056 79	53,44
GPT- 4Turbo_zer o_shot	0,284 09	0,115 389	0,339 69	- 1,070 27	0,320 904	0,3115 52	53,67
GPT- 4o_zero_sh ot	0,289 167	0,121 532	0,348 615	- 1,180 489	0,295 029	0,2954 76	57,58
GPT- 4_zero_shot	0,295 012	0,126 823	0,356 122	- 1,275 411	0,289 574	0,2842 44	58,10

Nota: Los resultados presentados en esta tabla muestran la precisión del modelo GPT-4 y sus sucesores GPT-4 Turbo y GPT-4o en la predicción de palabras complejas, evaluadas a través de métricas estadísticas como MAE, MSE, RMSE, R^2 , y correlaciones de Pearson y Spearman, y el porcentaje de coincidencia

permitiendo una comparación cuantitativa de su rendimiento en la tarea de simplificación del lenguaje.

Elaborado por: Scarlet Gutierrez e Ivette Zamora.

Evaluación de la predicción de la complejidad léxica entre modelos

Para la evaluación de los resultados se realizará una comparativa entre los diferentes modelos y las técnicas aplicadas. Estas combinaciones se detallan a continuación.

GPT-4, GPT-4 Turbo y GPT-4o con la técnica de few-shot learning.

Para estos modelos se les proporcionaron 20 ejemplos en el prompt para entrenarlos en la calificación de palabras identificadas como: 1. Moderately difficult, 2. Difficult y 3. Very difficult (se incluyeron 2 ejemplos con esta categoría), dando una sumatoria de 4 ejemplos por cada institución gubernamental pública.

En este sentido, como se observa en la tabla de resultados, la diferencia del MAE entre ambos modelos es pequeña pero significativa. Según (Tatachar, 2021) el MAE es una métrica que proporciona el promedio de la diferencia absoluta, esto hace que sea menos sensible a valores atípicos. Es decir, esta métrica nos permite evaluar la precisión de los modelos en el contexto de la identificación de palabras complejas, ya que permite entender mejor el error promedio de las predicciones sobre la complejidad léxica.

Con el objetivo de entender la diferencia del MAE entre ambos modelos, se calculará la diferencia porcentual utilizando la **Ecuación 11** Diferencia porcentual. Sabiendo que, el modelo GPT-4 tuvo un MAE de 0,24641, GPT-4 Turbo un MAE de 0.254051 y el modelo GPT-4o un MAE de 0.259385, se obtiene el siguiente resultado:

Entre GPT-4 y GPT-4 Turbo:

$$\% = \left(\frac{0.254051 - 0,24641}{0,24641} \right) \times 100 \approx 3.10\%$$

Entre GPT-4 y GPT-4o:

$$\% = \left(\frac{0.259385 - 0,24641}{0,24641} \right) \times 100 \approx 5.27\%$$

Como resultado, se tiene que el incremento porcentual en el MAE entre los modelos GPT-4 y GPT-4 Turbo es de aproximadamente 3.10% y para los modelos GPT-4 y GPT-4o es de 5.27%. Esto quiere decir que, GPT-4 realiza predicciones más precisas en

promedio en la identificación de palabras complejas en comparación con el GPT-4 Turbo, con respecto a la otra comparación el incremento es más notable, la diferencia en la precisión de GPT-4o podría indicar mejoras o variaciones en el entrenamiento. En resumen, GPT-4 Turbo y GPT-4o tienen un MAE mayor que GPT-4, pero con incrementos pequeños en ambos casos, siendo GPT-4o el que posee el mayor incremento. Con respecto al MSE o error cuadrático medio, el autor (Tatachar, 2021) menciona que, es la diferencia al cuadrado entre los valores reales y los valores predichos. Es decir, el MSE nos indica cuán cerca está la línea de mejor ajuste de un conjunto de puntos, se sabe que GPT-4 tiene un valor de 0,088819, GPT-4 Turbo tiene un valor de 0,091523 y GPT-4o un valor de 0,096301, con una diferencia porcentual de:

Entre GPT-4 y GPT-4 Turbo:

$$\% = \left(\frac{0.091523 - 0.088819}{0.088819} \right) \times 100 \approx 3.04\%$$

Entre GPT-4 y GPT-4o:

$$\% = \left(\frac{0.096301 - 0.088819}{0.088819} \right) \times 100 \approx 8.43\%$$

Esto quiere decir que la diferencia entre los valores reales y los valores predichos entre los modelos GPT-4 y GPT-4 Turbo es de 3.04%, con respecto a los modelos GPT-4 y GPT-4o es de 8.43%, lo que significa que el modelo GPT-4 Turbo es un 3.04% mayor que el de GPT-4, mientras que el MAE de GPT-4o es un 8.43% mayor que el de GPT-4. Es decir, GPT-4 tiene un rendimiento ligeramente mayor en términos de precisión y ajuste más cercano a los valores reales en comparación con GPT-4o y GPT-4 Turbo. Siguiendo con el RMSE que es la raíz cuadrada del error cuadrático medio MSE. Según el autor (Hodson, 2022) tomar la raíz no afecta los rangos relativos de los modelos, pero produce una métrica con las mismas unidades que (y), lo que representa convenientemente el error típico o 'estándar' para errores distribuidos normalmente. En este contexto, el modelo GPT-4 tiene un RMSE de 0,298025, GPT-4 Turbo tiene un RMSE de 0,302527 y GPT-4o un RMSE de 0,310325, con una diferencia porcentual de:

Entre GPT-4 y GPT-4 Turbo:

$$\% = \left(\frac{0.302527 - 0.298025}{0.298025} \right) \times 100 \approx 1.51\%$$

Entre GPT-4 y GPT-4o:

$$\% = \left(\frac{0.310325 - 0.298025}{0.298025} \right) \times 100 \approx 4.13\%$$

La diferencia porcentual entre los modelos GPT-4 y GPT-4 Turbo es de 1.51%, esto quiere decir que el modelo GPT-4 tiene un rendimiento ligeramente mayor que GPT-4 Turbo, con respecto a la otra evaluación, el modelo GPT-4 y GPT-4o tiene un incremento porcentual de 4.13%, es decir, el GPT-4o realiza predicciones menos precisas en comparación con GPT-4.

(Lewis-Beck & Skalaban, 1990), mencionan que El R² mide la proporción de varianza explicada por las variables independientes en un sentido estadístico. Sin embargo, esta medida no refleja la importancia de las variables en el modelo. Un R² de 1.00 no implica necesariamente que se haya encontrado una explicación válida para el fenómeno estudiado. En este sentido, el modelo GPT-4 tiene un valor R² de -0,593562, GPT-4 Turbo tiene un valor de -0,64207 y GPT-4o un valor de -0,727808, con una diferencia porcentual de:

Entre GPT-4 y GPT-4 Turbo:

$$\% = \left(\frac{-0.64207 - (-0.593562)}{0.593562} \right) \times 100 \approx 8.18\%$$

Entre GPT-4 y GPT-4o:

$$\% = \left(\frac{-0.727808 - (-0.593562)}{0.593562} \right) \times 100 \approx 22.61\%$$

Los valores en ambas comparaciones sugieren que los modelos tienen predicciones deficientes, y aunque los modelos no están ajustando bien los datos, el modelo GPT-4 Turbo es un 8.18% peor en términos de ajuste en comparación con el modelo GPT-4 y la diferencia de 22.61% nos indica que GPT-4o tiene un rendimiento significativamente peor que GPT-4.

Con respecto al valor de correlación de Pearson, los autores (Keskin & Topuz, 2021) mencionan que, el coeficiente de correlación de Pearson es una medida estadística que evalúa la fuerza y la dirección de la relación lineal entre dos variables cuantitativas. Su valor oscila entre -1 y 1, donde 1 indica una correlación positiva perfecta, -1 una correlación negativa perfecta y 0 indica que no hay correlación.

En este sentido, según la clasificación común de la magnitud de la correlación de la **¡Error! No se encuentra el origen de la referencia.**, el modelo GPT-4 tiene un valor de 0.07448, GPT-4 Turbo que tiene un valor de 0,31569 y el modelo GPT-4o con un valor de 0,316272 se encuentran en la categoría de correlación débil, que aunque existe una relación positiva entre las variables,

y el modelo GPT-4 Turbo tiene una correlación un poco más fuerte, todos los modelos tienen una capacidad débil para capturar la relación lineal entre las predicciones y los valores reales.

Por otra parte, con el coeficiente de correlación de Spearman, según indica (Keskin & Topuz, 2021) es una medida no paramétrica que evalúa la relación entre dos variables ordinales o continuas, sin asumir que la relación es lineal. A diferencia de Pearson, Spearman se basa en los rangos de los datos en lugar de los valores absolutos, lo que lo hace más robusto frente a outliers y distribuciones no normales. Su valor también varía entre -1 y 1, proporcionando una indicación de la dirección y la fuerza de la relación entre las variables.

Esto quiere decir, que los modelos GPT-4 con un valor de 0,056953, GPT-4 Turbo con un valor de 0,309387 y GPT-4o con un valor de 0,305679 tienen una correlación positiva, sin embargo, ya que estos valores son bajos, se considera una correlación débil. Esto significa que, aunque hay una tendencia a que a medida que una variable aumenta la otra también, esta relación no es lo suficientemente fuerte como para que las predicciones e los modelos se alineen de manera efectiva con los valores reales.

Por último, la coincidencia entre los modelos GPT-4, GPT-4 Turbo y GPT-4o, este último con una coincidencia de 53,44%, lo que indica, que GPT-4o tiene una mayor coincidencia en sus predicciones en comparación con GPT-4 y GPT-4 Turbo, es decir, GPT-4o es ligeramente más preciso en sus predicciones en relación a los valores reales en comparación con los otros dos modelos.

GPT-4 con la técnica de zero-shot learning, GPT-4 Turbo con la técnica de few-shot learning y GPT-4o con la técnica de zero-shot learning.

Para la comparación entre GPT-4 con la técnica zero-shot learning, GPT-4 Turbo con la técnica de few-shot learning y GPT-4o con la técnica de zero-shot learning, se pueden observar los siguientes resultados. Para el MAE de GPT-4 (zero-shot) el valor es de 0.295012, para GPT-4 Turbo (few-shot) su valor es de 0.254051 y para GPT-4o (zero-shot) el valor es de 0.289167l. La diferencia porcentual entre los modelos GPT-4 y GPT-4 Turbo es de 16.14% lo que indica que GPT-4 Turbo con la técnica few-shot tiene un rendimiento mejor que GPT-4 con la técnica zero-shot en términos de MAE y GPT-4 con GPT-4o es de 2.02%, esto quiere decir que GPT-4o usando la técnica zero-shot tiene un rendimiento ligeramente mejor que GPT-4.

Entre GPT-4 (zero-shot) y GPT-4 Turbo (few-shot):

$$\% = \left(\frac{0.295012 - 0.254051}{0.254051} \right) \times 100 \approx 16.14\%$$

Entre GPT-4 (zero-shot) y GPT-4o (zero-shot):

$$\% = \left(\frac{0.295012 - 0.2891671}{0.2891671} \right) \times 100 \approx 2.02\%$$

Con respecto al MSE, el modelo GPT-4 (zero-shot) tuvo un valor de 0.126823, GPT-4 Turbo (few-shot) tuvo un valor de 0.091523 y el modelo GPT-4o (zero-shot) tuvo un valor de 0.121532. La diferencia porcentual entre los modelos GPT-4 y GPT-4 Turbo es de 38.53%, lo que significa que el modelo GPT-4 Turbo (few-shot) tiene un rendimiento significativamente mayor en términos de precisión y ajuste más cercano a los valores reales en comparación con GPT-4 (zero-shot), con los modelos GPT-4 (zero-shot) y GPT-4o (zero-shot) es de 4.36%, esto quiere decir que GPT-4o tiene un rendimiento ligeramente mejor que GPT-4.

Entre GPT-4 (zero-shot) y GPT-4 Turbo (few-shot):

$$\% = \left(\frac{0.126823 - 0.091523}{0.091523} \right) \times 100 \approx 38.53\%$$

Entre GPT-4 (zero-shot) y GPT-4o (zero-shot):

$$\% = \left(\frac{0.126823 - 0.121532}{0.121532} \right) \times 100 \approx 4.36\%$$

Para el R2, entre GPT-4 (zero-shot) y GPT-4 Turbo (few-shot) se tiene una diferencia porcentual de 98.59% lo que indica, que el modelo GPT-4 Turbo (few-shot) tienen un ajuste significativamente mejor que el modelo GPT-4 (zero-shot), esto significa que, aunque ambos modelos tienen valores negativos, el modelo GPT-4 Turbo es mejor en su capacidad de predicción. Para los modelos GPT-4 (zero-shot) y GPT-4o (zero-shot) el porcentaje es de 8.04%, lo que quiere decir que GPT-4o tiene un rendimiento ligeramente mejor en la predicción y valores reales que el modelo GPT-4.

Entre GPT-4 (zero-shot) y GPT-4 Turbo (few-shot):

$$\% = \left(\frac{-1.275411 - (-0.64207)}{-0.64207} \right) \times 100 \approx 98.59\%$$

Entre GPT-4 (zero-shot) y GPT-4o (zero-shot):

$$\% = \left(\frac{-1.275411 - (-1.180489)}{-1.180489} \right) \times 100 \approx 8.04\%$$

En cuanto a la correlación de Pearson, los tres modelos presentan una correlación débil, ya que los valores están por debajo del 0.5, esto indica que, aunque existe una relación positiva entre las predicciones de los modelos y los valores reales, no es lo suficientemente fuerte como para garantizar que las predicciones sean precisas. Aunque los valores son muy similares, GPT-4 (zero-shot) con una correlación de 0.289574 es ligeramente mejor en términos de predicción con los valores reales en comparación con GPT-4o (zero-shot) y GPT-4 Turbo (few-shot).

Con respecto a la correlación de Spearman, el modelo GPT-4 tiene un valor de 0.284244, GPT-4 Turbo (few-shot) es de 0.309387 y con respecto a GPT-4o (zero-shot) es un valor de 0.295476, al igual que Pearson estos valores son bajos, es decir, tienen una correlación débil. Lo que quiere decir que, aunque hay una tendencia de que a medida que una variable aumenta, la otra también, la relación no es lo suficientemente fuerte como para que las predicciones de los modelos se alineen con los valores reales.

Por último, la coincidencia del modelo GPT-4 (zero-shot) es de 58.10%, esto indica que tiene una mayor coincidencia en sus predicciones en comparación con el modelo GPT-4o (zero-shot) con una coincidencia del 57.58% y GPT-4 Turbo (few-shot) que tiene una coincidencia del 51.67%. Es decir, el modelo GPT-4 (zero-shot) es ligeramente más preciso en sus predicciones en relación con los valores reales en comparación con el modelo GPT-4o (few-shot) y GPT-4 Turbo (few-shot).

GPT-4, GPT-4 Turbo y GPT-4o con la técnica de zero-shot learning.

Para la técnica de zero-shot learning, los modelos no reciben ningún ejemplo en las instrucciones, sino que deben generar respuestas sin información adicional. De esta manera, el resultado de una de las medidas estadísticas MAE a simple vista presenta un error absoluto ligeramente menor, recordemos que esta medida según menciona (Camarillo-Peñaranda, 2013) mide la magnitud promedio del error entre los datos medidos y los datos estimados por el modelo, y mientras más cercana este al cero es mejor, ya que sugiere un buen rendimiento.

Para clarificar el porcentaje de diferencia relativa entre los modelos se calcula la diferencia porcentual mediante la fórmula **Ecuación 11** Diferencia porcentual que se menciona en la metodología, teniendo como resultado:

Siendo 0.289167 el valor de MAE del modelo GPT-4 Turbo y 0.28409 el valor de MAE del modelo GPT-4o.

$$\% = \left(\frac{0.289167 - 0.28409}{0.28409} \right) \times 100 \approx 1.786\%$$

El resultado muestra que el modelo GPT-4o presenta un MAE aproximadamente de 1.786% mayor a GPT-4 Turbo indicando que el modelo Turbo es más preciso en términos de error absoluto medio, aunque su diferencia sigue siendo pequeña. Para el valor del MAE, GPT-4 obtuvo un valor de 0,295012, mientras que el sucesor GPT-4 Turbo obtuvo 0,28409.

$$\% = \left(\frac{0.295012 - 0.28409}{0.295012} \right) \times 100 \approx 3.70\%$$

El resultado muestra que el modelo GPT-4 presenta un porcentaje de 3.70% mayor que GPT-4 Turbo. El valor del MAE para el modelo GPT-4 es de 0,295012 y el del modelo GPT-4o es de 0,289167, siendo su diferencia porcentual de:

$$\% = \left(\frac{0.295012 - 0.289167}{0.295012} \right) \times 100 \approx 1.98\%$$

Como resultado, la diferencia entre ambos modelos es de 1.98% siendo GPT-4o mejor que GPT-4. En general el modelo con menor MAE es el modelo GPT-4 Turbo, lo que indica que sus predicciones son más precisas que las de GPT-4 y GPT-4o. En el caso de la segunda medida calculada, MSE, como lo menciona (Na & Neuhoff, 2001) es simplemente la diferencia que existe entre el valor que nuestro modelo ha predicho y el valor real de la observación, en este caso del corpus. En el modelo de GPT-4 Turbo el valor es de 0,115389 y en el modelo de GPT-4o es de 0,121532, siendo su diferencia porcentual de:

$$\% = \left(\frac{0.121532 - 0.115389}{0.115389} \right) \times 100 \approx 5.319\%$$

Como resultado se puede interpretar que GPT-4o tiene un MSE aproximadamente de 5.319% mayor que GPT-4 Turbo, lo que señala a GPT-4 Turbo más eficiente en términos de reducir errores grandes en comparación con GPT-4o. Para el modelo GPT-4 el valor de MSE es de 0,126823, siendo así que el modelo sucesor GPT-4 Turbo sigue siendo líder en cuanto a reducir grandes errores. Para el valor de RMSE los autores (Soto-Bravo & González-Lutz, 2019) señalan que esta medida no es útil por sí mismo para calificar la confiabilidad del método de simulación, pero así mismo indican que sirve para medir, en promedio, cuánto se

alejan los datos observados de los estimados por el modelo de regresión, lo cual sirve para saber información acerca del desempeño del modelo.

El valor arrojado de RMSE en la ejecución del modelo GPT-4 Turbo es 0,33969 y para el modelo GPT-4o es de 0,348615. El resultado del cálculo de la diferencia porcentual es:

$$\% = \left(\frac{0.348615 - 0.33969}{0.33969} \right) \times 100 \approx 2.625\%$$

El cálculo indica que GPT-4o tiene un RMSE aproximadamente de 2.625% mayor que GPT-4 Turbo, resultando que GPT-4 Turbo maneja mejor la dispersión de los errores, a su vez indica un mejor ajuste general en las predicciones, es decir, sus predicciones están más cerca de los valores reales en comparación con GPT-4o, lo que puede ser relevante en aplicaciones donde la precisión es crucial.

El resultado del RMSE del modelo GPT-4 es de 0,356122, este valor es mayor en comparación con GPT-4o y GPT-4 Turbo, siendo este último el más destacado por su valor menor.

En cuanto al valor de R2, el modelo de GPT-4 Turbo dio como resultado el valor de -1.07027, mientras que con el otro modelo sucesor tiene un valor de -1.18049, en cuanto a su diferencia relativa se tiene que su resultado es:

$$\% = \left(\frac{-1.18049 - (-1.07027)}{-1.07027} \right) \times 100 \approx -10.31\%$$

GPT-4o tiene un R2 aproximadamente un 10.31% más bajo, más negativo que GPT-4 Turbo, esto quiere decir que, el primer modelo ajusta peor los datos que el segundo, aunque ambos modelos tienen un rendimiento deficiente en este caso. Para GPT-4 el valor obtenido es de -1,275411.

Para la autora (Palma Roxanna Patricia., 2022) "cuando el valor de R2 es igual a +1 existe una relación lineal perfecta entre x e y , es decir, el 100 % del cambio en y se explica por el cambio en x . Cuando es $0 < R2 < 1$, existe una relación lineal más débil entre x e y , es decir, no toda la variación de y se explica por la variación de x ". Teniendo en cuenta lo mencionado por la autora los modelos tienen una relación débil en su capacidad para explicar la variación en los datos, ya que sus valores de R2 están por debajo de 1, lo cual señala que los modelos ajustan mal los datos.

En cuanto a la medida de Pearson, según (Ortiz Pinilla et al., 2021), si sus valores extremos son (± 1) están estrechamente

asociados con los mejores ajustes posibles, que, de acuerdo con el criterio de mínimos cuadrados, corresponden a líneas rectas. Esto implica que una correlación de -1 o $+1$ representa una relación perfecta y lineal entre las variables analizadas. Lo que no sucede con los valores que arrojan ambos modelos, ya que los dos indican una correlación positiva débil entre las predicciones del modelo y los valores reales del corpus.

Aunque el modelo de GPT-4 Turbo tiene una correlación un poco más fuerte, los valores sugieren que los tres modelos tienen una capacidad limitada para capturar la relación lineal entre las predicciones y los valores reales del corpus.

En cambio, en el caso de Spearman que es exactamente el mismo que el coeficiente de correlación de Pearson, solo que es calculado sobre el rango de observaciones, así que sus valores se interpretan igual que el de Pearson (Martínez Ortega et al., 2009). El resultado de Spearman de los modelos presenta una relación positiva débil entre las predicciones y los valores reales, lo que sugiere que, aunque hay una tendencia general a que las predicciones se alineen con los valores reales, la relación no es muy fuerte.

Por último, el campo de coincidencias sugiere que a pesar de que en las medidas el modelo GPT-4 Turbo fue un poco superior al modelo de GPT-4o y al GPT-4, este último obtuvo la tasa de coincidencias más alta con una puntuación de 58,10% de coincidencias entre los valores generados por él y los valores proporcionados por el corpus en relación al campo de complejidad del token, siendo así que este modelo demuestra ser ligeramente superior para esta tarea con relación a los otros modelos.

GPT-4 con la técnica de few-shot learning , GPT-4o con la técnica de zero-shot learning y GPT-4o con la técnica de few-shot learning.

Los resultados de GPT-4o con ambas técnicas son esenciales en la evaluación de estos y sirven para comprender cómo el modelo adapta su rendimiento en tareas de predicción de complejidad léxica, dependiendo de la disponibilidad de ejemplos previos en el entrenamiento. Además, se añade a esta evaluación el modelo GPT-4 con la técnica de few-shot learning.

Para GPT-4o el valor de MAE aplicando la técnica de few-shot es de 0.259385, en cambio para zero-shot es de 0.289167. Para comprender mejor el resultado se calcula la diferencia porcentual a continuación:

$$\% = \left(\frac{0.289167 - 0.259385}{0.289167} \right) \times 100 \approx 10.30\%$$

La diferencia porcentual es de 10.30%, lo que apunta que el MAE de GPT-4o con few-shot es un 10.30% menor que el mismo modelo con zero-shot, siendo así que hay una mejora en la precisión de las predicciones cuando se utiliza el enfoque few-shot. En cambio, el resultado del modelo GPT-4 es de 0,24641, siendo así que destaca entre su sucesor con ambas técnicas.

Recordemos que MAE como se mencionó anteriormente es una medida que mientras más cercana es al cero mejor es su desempeño, el cual es el caso del modelo GPT-4 usando la técnica de few-shot y esto se puede constatar solo al comparar los resultados.

Por otro parte, el resultado de la medida estadística MSE en el caso de la técnica de zero-shot con GPT-4o el valor es de 0.121532, mientras que para la técnica con ejemplos (GPT-4o con few-shot) es de 0.096301, y la diferencia porcentual es de:

$$\% = \left(\frac{0.121532 - 0.096301}{0.121532} \right) \times 100 \approx 20.76\%$$

Al igual que la medida de MAE, en MSE si el resultado se acerca al cero indica un mejor desempeño, a su vez la diferencia porcentual indica que GPT-4o con few-shot es un 20.76% menor que zero-shot, mientras que el resultado de GPT-4 es de 0,088819. En términos prácticos, el enfoque de few-shot para el modelo GPT-4 reduce significativamente los errores en comparación con su modelo sucesor tanto con la técnica zero-shot como con few-shot, lo que resalta que GPT-4 con few-shot es más eficaz para minimizar tanto errores pequeños como grandes.

Para RMSE en la técnica sin ejemplos GPT-4o (zero-shot) arrojo un resultado de 0,348615, este valor refleja la dispersión promedio de los errores cometidos por el modelo en comparación con los valores del corpus. El termino de dispersión se refiere a la medida en cuánto varían los datos entre sí, en relación con un valor central, es decir, si los datos están más cerca o más lejos de ese valor central. A mayor dispersión, más alejados están los datos entre sí, en cambio a menor dispersión, más concentrados están alrededor del valor central. (Mayorga-Ponce et al., 2021)

Por el contrario, para GPT-4o few-shot el resultado es de 0,3110325 siendo así que, en promedio, el modelo está cometiendo errores más pequeños en comparación con la otra técnica, este a su vez, es un valor más bajo, lo que sugiere que las predicciones son más precisas.

Para GPT-4 el valor de RSME es de 0,298025, este valor nos indica que el modelo destaca en cometer errores más pequeños que su modelo sucesor.

En R2 los valores de GPT-4o con ambas técnicas son negativos lo que significa que sus predicciones en cuanto a la complejidad léxica no es muy eficiente. Cabe destacar que GPT-4o con few-shot el valor de R2 es ligeramente menos negativo que con zero-shot, es decir, el rendimiento mejora un poco, pero el mejor valor lo tiene el modelo de GPT-4 con -0,593562, a pesar de también ser un valor negativo, es mucho menor que los otros resultados. Por otra parte, con Pearson el valor del modelo GPT-4o con few-shot dio como resultado 0.316272, lo cual indica que existe una correlación positiva débil entre las predicciones del modelo y los valores reales. Esto significa que hay una relación lineal moderadamente débil entre las predicciones y los valores proporcionados en el corpus.

En cambio, el valor del modelo GPT-4o con zero-shot a pesar de que también indica una correlación positiva débil, es un poco más bajo que few-shot, lo que señala que GPT-4o con zero-shot es menos confiable y coherente con los valores reales proporcionados por el corpus.

El valor del Pearson con el modelo GPT-4 es de 0,07448 lo que indica que su correlación es casi nula en esta tarea, ya que su valor es más cercano al cero, siendo así que entre los tres el modelo GPT-4o con few-shot permite capturar mejor la relación entre las variables.

Para Spearman el valor de few-shot para el modelo de GPT-4o es de 0,309387, GPT-4o con zero-shot el valor es de 0,295476, y para GPT-4 es de 0,056953, estos valores indican que existe una correlación monótona positiva débil, según lo mencionan (Hauke, Tomasz Kossowski, 2011) la correlación monótona positiva débil hace alusión a la relación que existe entre dos variables, en donde ambas tienden a aumentar, pero no necesariamente a un ritmo constante. Teniendo en cuenta lo mencionado, podemos decir que el modelo GPT-4o con ambas técnicas, aunque presentan un desempeño similar, no logran tener una relación robusta entre las variables. En cambio, GPT-4 tiene un valor super bajo lo que indica que tiene el peor desempeño, mostrando que su correlación es casi inexistente.

Finalmente, el porcentaje de coincidencias para GPT-4 con few-shot es de 37,59% , para GPT-4o con few-shot es de 51,67% y para zero-shot es de 57,58%. A pesar de que en las medidas anteriores few-shot destaque por ser ligeramente mejor que zero-shot para las predicciones, vemos que esta última tiene un mayor porcentaje de coincidencias con el corpus, según (Brown

et al., 2020) esto se debe a que " Aunque el aprendizaje de few-shot permite que el modelo se adapte más efectivamente a una tarea específica con ejemplos limitados, el aprendizaje de zero-shot a veces puede resultar en porcentajes de coincidencia exacta más altos porque el modelo se ve obligado a generalizar en función del conocimiento preexistente de gran escala, datos de entrenamiento, lo que lleva a respuestas más diversas, pero menos precisas".

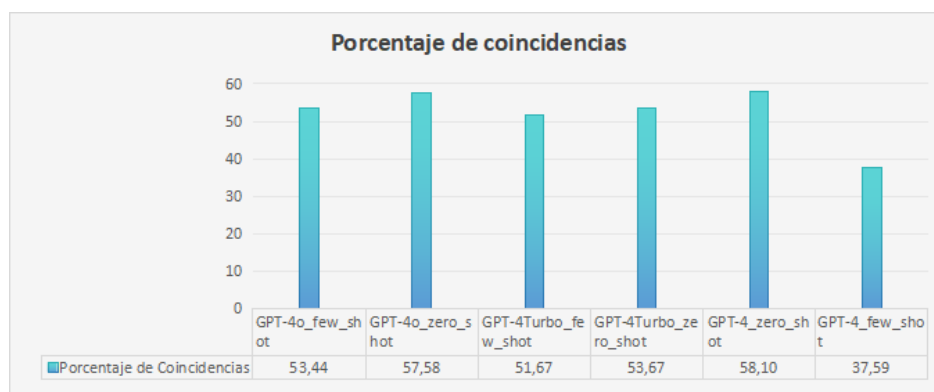
Evaluación del porcentaje de coincidencia entre los modelos.

En cuanto a los resultados de porcentaje de coincidencias entre los modelos GPT-4, GPT-4 Turbo con few-shot y zero-shot, y GPT-4o con las mismas técnicas de aprendizaje, se presentan en el siguiente gráfico donde se puede visualizar el resultado de cada ejecución:

Figura

5

Porcentaje de coincidencias de los modelos



Nota: Resultados de los porcentajes de coincidencias de las ejecuciones realizadas en ambos modelos con las técnicas de few-shot y zero-shot. **Elaborado por:** Scarlet Gutierrez e Ivette Zamora.

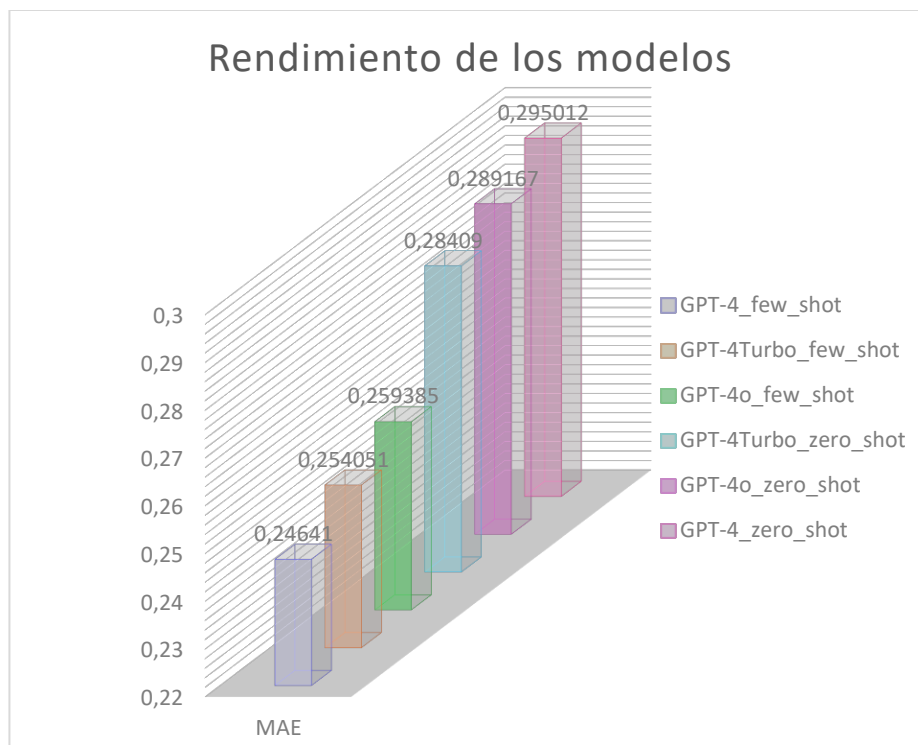
Como se puede visualizar en el gráfico anterior, GPT-4 utilizando zero-shot learning es el modelo con más coincidencias con el corpus a diferencia de los otros modelos, con el 58.10% de coincidencias. Aunque las otras ejecuciones no están tan alejadas de este resultado, presentan un desempeño ligeramente inferior, destacando que GPT-4o con zero-shot alcanzó un 57.58%, mientras que las demás oscilaron entre el 51%, 53% y el más bajo fue el del modelo GPT-4 con la técnica de few-shot. Esto sugiere que la técnica de zero-shot learning en GPT-4o ofrece una mayor precisión.

Rendimiento de los modelos en función del MAE

Figura

6

Rendimiento de los modelos en función del MAE



Nota: Resultados del rendimiento del modelo GPT-4 y sus sucesores en función al MAE con las técnicas de few-shot y zero-shot. **Elaborado por:** Scarlet Gutierrez e Ivette Zamora.

Con respecto al rendimiento de los modelos en función del MAE se tiene que, el modelo GPT-4_zero_shot (0.295012) presenta el MAE más alto de todos los modelos, esto significa que las predicciones realizadas por este modelo tienen una mayor diferencia a los valores reales.

Por otro lado, el modelo GPT-4_few_shot (0.24641) tiene un MAE más bajo que el resto de los modelos, esto significa que es más preciso en la predicción de la complejidad léxica.

El modelo GPT-4Turbo_few_shot tiene un MAE de (0.254051) el cual, es intermedio entre los otros dos modelos. Aunque no es el más preciso, su rendimiento es mejor que el de GPT-4o_few_shot (0.259385), esto indica que el modelo Turbo tiene un desempeño aceptable en la predicción de la complejidad léxica. Sin embargo, su MAE es más alto que el de GPT-4_few_shot (0.24641), lo que indica que, aunque tiene una mejora

Con respecto al modelo GPT-4o_zero_shot (0.289167) el rendimiento es superior al modelo GPT-4_zero_shot (0.295012) ya que tiene un MAE ligeramente más bajo, sin embargo, el

modelo GPT-4 Turbo_zero_shot sigue siendo el modelo con el mejor desempeño en términos de precisión, con el MAE más bajo de todos los modelos con la técnica zero-shot.

Por último, el modelo GPT-4o_few_shot (0.259385) posee el MAE más alto de todos los modelos, esto indica que el modelo con esta técnica tiene un desempeño peor que los otros modelos, lo que nos indica también que la técnica few-shot no siempre mejora la precisión.

Resumen final

- GPT-4_few_shot (0.24641): Mejor precisión en general.
- GPT-4Turbo_few_shot (0.254051): Rendimiento intermedio, superando a GPT-4o_few_shot.
- GPT-4o_few_shot (0.259385): Aunque utiliza la técnica few-shot, tiene el MAE más alto entre los modelos con esta técnica.
- GPT-4Turbo_zero_shot (0.28409): Mejor rendimiento en la categoría zero-shot.
- GPT-4o_zero_shot (0.289167): Supera ligeramente a GPT-4_zero_shot.
- GPT-4_zero_shot (0.295012): Menor precisión en general.

En resumen, los primeros tres lugares con el MAE más bajo lo ocupan los tres modelos con la técnica few-shot learning, lo que nos indica que esta técnica es más efectiva en la mejora de la precisión de los modelos en comparación con zero-shot learning. Es decir, el aprendizaje con ejemplos adicionales (few-shot) mejora la precisión del modelo. Sin embargo, GPT-4o_few_shot es una excepción, ya que tiene un MAE más alto que otros modelos con la misma técnica, lo que sugiere que la mejora no es automática y depende del modelo específico.

En la categoría zero-shot learning, GPT-4Turbo_zero_shot es el modelo con mejor rendimiento, superando a los modelos GPT-4o_zero_shot y GPT-4_zero_shot. Esto quiere decir que, aunque los resultados de estos modelos usando la técnica de zero-shot learning no superen la precisión obtenida de los modelos con la técnica few-shot learning es una buena alternativa para casos donde no se tenga ejemplos previos para proporcionar al modelo.

CAPÍTULO IV

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

1. En cuanto a las características lingüísticas se pudo encontrar un total de 40 características que han sido calculadas y esto permitirá el correspondiente análisis para trabajos futuros en cuanto a la complejidad léxica.
2. Como segunda conclusión se tiene que existe una diferencia notable entre los modelos GPT-4, GPT-4o y GPT-4 Turbo al aplicar las técnicas de zero-shot y few shot, siendo la más destacada la técnica de zero-shot learning especialmente en el modelo de GPT-4 con un porcentaje de coincidencias con el corpus superior a los demás modelos.
3. Con respecto al rendimiento de los modelos en función del MAE se concluye que, el modelo GPT-4 con zero-shot (0.295012) presenta el MAE más alto de todos los modelos, mientras que el modelo GPT-4 con few-shot (0.24641) tiene el MAE más bajo, esto quiere decir que las predicciones realizadas por este modelo con la técnica few-shot se acercan más a los valores reales. Es decir, la técnica few-shot learning es más efectiva en la mejora de la precisión de los modelos que la técnica zero-shot learning.
4. Los modelos en general presentan una tendencia entre sus porcentajes de coincidencias y entre las demás métricas estadísticas, siendo así que GPT-4o y GPT-4 Turbo presentan resultados similares en ambas técnicas, con diferencias menores a 2% en few-shot y 3,91% en zero-shot, mientras que los demás oscilan entre el 50% para arriba, diferenciándose el modelo GPT-4 con few shot con un porcentaje super menor.
5. Los modelos GPT-4 y sus sucesores han demostrado cierta capacidad para predecir la complejidad léxica en textos públicos gubernamentales, aunque con variaciones en su desempeño según la técnica utilizada. Si bien los modelos muestran cierto grado de efectividad, los porcentajes obtenidos indican que aún existen limitaciones en su precisión, por lo que su uso en aplicaciones reales requeriría refinamientos adicionales.
6. Los resultados obtenidos sugieren que los modelos GPT-4 y sus sucesores tienen la capacidad de identificar

y predecir la complejidad léxica en documentos públicos, pero no necesariamente de reducirla de manera eficaz y óptima para incrementar la comprensión ciudadana.

7. La implementación de modelos GPT-4 y sus sucesores en la redacción de textos públicos tienen potencial de mejorar la transparencia y accesibilidad a la información gubernamental, ya que pueden identificar y analizar la complejidad léxica de los documentos. Sin embargo, los resultados obtenidos muestran que, si bien estos modelos pueden predecir la complejidad con cierto grado de precisión, su desempeño no es completamente uniforme en las dos técnicas aplicadas.
8. Finalmente, para concluir, se puede decir que, aunque puede ser factible desarrollar una herramienta basada en GPT-4 para simplificar documentos públicos, su implementación efectiva requiere una combinación de estos modelos de IA y la revisión humana para asegurar que la información sea accesible sin comprometer su exactitud y relevancia.

Recomendaciones

1. Se recomienda abastecer el corpus con más textos de documentos públicos de diversas áreas gubernamentales de Guayaquil, asegurando una mayor representatividad y diversidad lingüística. Además, es importante incluir textos con distintos niveles de complejidad léxica para mejorar la capacidad de generalización de los modelos.
2. Para la técnica de few shot se recomienda el uso de ejemplos representativos y bien estructurados que ayuden al modelo a comprender mejor los patrones de complejidad léxica en los textos públicos. Es fundamental seleccionar ejemplos con diversos niveles de dificultad léxica, asegurando que el modelo pueda captar diferencias sutiles en la estructura y el uso del lenguaje.

Trabajos futuros

En el contexto de la evaluación del desempeño de los sucesores del modelo Generative Pre-trained Transformer GPT-4 en la predicción de la complejidad léxica en textos públicos, se han identificado diversas áreas que requieren atención y desarrollo adicional. A pesar de los avances logrados en la comprensión de la complejidad léxica mediante el uso de modelos de aprendizaje profundo, aún existen oportunidades para mejorar la accesibilidad y la claridad de la información pública.

A continuación, se presentan algunas propuestas para trabajos futuros que podrían enriquecer la investigación en este campo:

1. Ampliación del Corpus de datos: Se sugiere la inclusión de más registros en el corpus para abarcar más entidades gubernamentales, así como textos de diferentes contextos y niveles de complejidad. Esta inclusión permitirá una evaluación más robusta en el rendimiento de los modelos en la predicción de la complejidad léxica.
2. Exploración de otros modelos de aprendizaje profundo: Aunque el enfoque actual se centra en los sucesores del modelo GPT-4, en futuros trabajos podrían investigar el rendimiento de otros modelos de aprendizaje profundo, para comparar su eficacia en la identificación de la complejidad léxica en textos públicos.
3. Ampliar la investigación a un contexto multilingüe: Adaptando y entrenando los modelos en diversos idiomas resultaría importante para la administración

pública en áreas donde se utilizan múltiples idiomas oficiales, lo que permitirá realizar un análisis más inclusivo y exhaustivo de la complejidad textual.

4. Ampliar el conjunto de características lingüísticas: Investigar y calcular nuevas características lingüísticas permitirá un análisis más profundo y detallado de los textos, mejorando de esta manera la precisión en la evaluación de la complejidad léxica.
5. Estos trabajos futuros presentan oportunidades de mejora en cuanto a la predicción de la complejidad textual y su uso en diversos ámbitos, lo que permitirá mejorar el desempeño de los modelos actuales y expandir su aplicación en nuevas áreas de estudio y práctica. Esto no solo facilitará una mejor comprensión de los textos públicos, sino que también promoverá una mayor participación ciudadana y reforzará la transparencia en la comunicación del gobierno.

REFERENCIAS BIBLIOGRÁFICAS

- Alejandro Ayuso Expósito. (2024). *Lectura Fácil: Aplicación web para evaluar adaptaciones automáticas*.
- Alexandra Ortiz-Zambrano, J., & Montejo-Raéz, A. (2020). *Barriers in Reading Comprehension of University Students: Analysis of the Complicated Words Annotated in the VYTEDU-CW Corpus*. 10(5).
- Alfonso-Robaina, D. I., Villazón-Gómez, A. I., Enrique Milanes-Amador, P. I., Rodríguez-González III, A., & Espín-Alonso, R. I. (2011). *PROCEDIMIENTO GENERAL DE REDISEÑO ORGANIZACIONAL PARA MEJORAR EL ENFOQUE A PROCESOS / GENERAL PROCEDURE OF ORGANIZATIONAL REDESIGN TO IMPROVE THE PROCESS APPROACH*.
- Alizadeh, M., Kubli, M., Samei, Z., Dehghani, S., Zahedivafa, M., Bermeo, J. D., Korobeynikova, M., & Gilardi, F. (2023). *Open-Source LLMs for Text Annotation: A Practical Guide for Model Setting and Fine-Tuning*. <http://arxiv.org/abs/2307.02179>
- Alliende, F. (2002). *La legibilidad de los textos: manual para la evaluación, selección y elaboración de textos* (A. Bello, Ed.; 2nd ed.).
- Álvarez, C., Vega, M. De, Psicothema, M. C.-, & 1998, undefined. (1998). La sílaba como unidad de activación léxica en la lectura de palabras trisílabas. *Redalyc.Org*. <https://www.redalyc.org/pdf/727/72710211.pdf>
- Anderson, J. (1983). Lix and Rix: Variations on a Little-known Readability Index. *Journal of Reading*, Vol. 26(No. 6), 1-496.
- Angamarca, K. (2023). *Estudio comparativo de herramientas de procesamiento de lenguaje natural open source y software propietario para el desarrollo de aplicaciones basadas en inteligencia artificial*.
- Armas, A., Tutor, B., Carlos, :, & Álvarez González, J. (2016). *LA INFLUENCIA DE LA LONGITUD DE LA PALABRA EN EL RECUERDO DE PALABRAS VISUALMENTE PRESENTADAS*.
- Baktash, J. A., & Dawodi, M. (2023). Gpt-4: A Review on Advancements and Opportunities in Natural Language Processing. *Journal of Electrical Electronics Engineering*, 2(4). <https://doi.org/10.33140/jeee.02.04.19>
- Baquadano, M. M. (2006). *Legibilidad y variabilidad de los textos*.
- Barrio Cantalejo, I., & Simón Lorda, P. (2003). Medición de la legibilidad de textos escritos. Correlación entre método manual de Flesch y métodos informáticos. *Atención Primaria*, 31(2), 104-108. [https://doi.org/10.1016/S0212-6567\(03\)79146-X](https://doi.org/10.1016/S0212-6567(03)79146-X)
- Barton, J. J. S., Hanif, H. M., Eklinder Björnström, L., & Hills, C. (2014). The word-length effect in reading: A review. *Cognitive Neuropsychology*, 31(5-6), 378-412. <https://doi.org/10.1080/02643294.2014.895314>

- Bautista, S., Drndarević, B. D., Hervás, R., Saggion, H., & Gervás, P. (2012). Análisis de la Simplificación de Expresiones Numéricas en Español mediante un Estudio Empírico. *Linguamática*, 4(2), 27-41.
- Belloch, C. (2011). Las tecnologías de la información y comunicación (T.I.C.). *Universidad de Valencia*, 1.
- Bono Cabré, R. (2012). Diseños cuasi-experimentales y longitudinales. *OMADO (Objectes i MAterials DOcents)*. <https://diposit.ub.edu/dspace/handle/2445/30783>
- Borjabad Salud. (2023). *La intersección entre la adquisición del lenguaje y la inteligencia artificial: explorando el potencial de los modelos de lenguaje natural*.
- Bosque, I. (2016). *EL SUSTANTIVO*. 2, 111-123.
- Botchkarev, A. (2018). *Evaluating performance of regression machine learning models using multiple error metrics in Azure Machine Learning Studio*. <https://ssrn.com/abstract=3177507>
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). *Language Models are Few-Shot Learners*. <https://commoncrawl.org/the-data/>
- Campos Daniela, Contreras Paula, Riffo Bernardo, & Véliz Mónica. (2013). *Vista de Complejidad textual, lecturabilidad y rendimiento lector en una prueba de comprensión en escolares adolescentes*.
- Campos Saavedra, D., Contreras Carmona, P., Riffo Ocares, B., Véliz, M., & Reyes Reyes, A. (2014). Complejidad textual, lecturabilidad y rendimiento lector en una prueba de comprensión en escolares adolescentes. *Universitas Psychologica*, 13(3). <https://doi.org/10.11144/javeriana.upsy13-3.ctrl>
- Cesteros, J. S. (2023). *Aproximaciones a la simplificación léxica mediante aprendizaje profundo Trabajo Fin de Máster*.
- Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7, 1-24. <https://doi.org/10.7717/PEERJ-CS.623>
- Choi, H.-S., Trivedi, P., Constant, M., Fort, K., & Guillaume, B. (2024). *Beyond Model Performance: Can Link Prediction Enrich French Lexical Graphs?* <https://doi.org/10.13039/501100001665>
- Dapena, J. A. Porto. (1986). *Los pronombres*. 256. https://books.google.com/books/about/Los_pronombres.html?hl=es&id=uUxdAAAAMAAJ
- De Miguel, E. (2004). *QUÉ SIGNIFICAN ASPECTUALMENTE ALGUNOS VERBOS Y QUÉ PUEDEN LLEGAR A SIGNIFICAR**.
- De, U. T., De Lectura, A., Pablo Barreyro, J., Burin, D. I., & Aníbal Duarte, Y. D. (2009). Capacidad de la memoria de trabajo verbal: Validez y fiabilidad de una tarea de amplitud de lectura. *Interdisciplinaria*,

- 26(2), 207-228.
https://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1668-70272009000200003&lng=es&nrm=iso&tlng=pt
- Desai, A., North, K., Zampieri, M., & Homan, C. M. (2021). *LCP-RIT at SemEval-2021 Task 1: Exploring Linguistic Features for Lexical Complexity Prediction*.
<https://sites.google.com/view/lcpsharedtask2021/hom>
- Dhiman Devansh. (2023). *El Futuro de la IA Conversacional: Explorando los Últimos Avances en Chat GPT-4* (D. Devansh, Ed.).
- Díaz, G. B. (2024). *TFG_GUILLERMO_DIAZ_BENITO*.
- Díaz Rojas, A. (2005). De los documentos públicos. (*Doctoral Dissertation*)
- El Universo. (2024). *En plena era digital, en Ecuador 472.228 personas adultas no saben leer ni escribir: conoce las provincias que tienen la mayor tasa de analfabetismo*.
<https://www.eluniverso.com/noticias/informes/en-plena-era-digital-en-ecuador-472228-personas-adultas-no-saben-leer-ni-escribir-conoce-las-provincias-que-tienen-la-mayor-tasa-de-analfabetismo-nota/>
- Farrús, M., Anguita, J., Hernando, J., & Cerdà, R. (2005). *Fusión de sistemas de reconocimiento basados en características de alto y bajo nivel*.
- Fellbaum, C. (2010). WordNet. *Theory and Applications of Ontology: Computer Applications*, 231-243. https://doi.org/10.1007/978-90-481-8847-5_10
- Fernández Álvarez, F. (2024). *MetaCity text-to-KPI: Sistema de Extracción de Indicadores Urbanos por Procesamiento de Lenguaje Natural*.
- Fuentes-Monsálves, L. I. (2009). Diagnóstico de comprensión lectora en educación básica en Villarrica y Loncoche, Chile. *Perfiles Educativos*, XXXI (125), 23-37.
<http://www.redalyc.org/articulo.oa?id=13211980003>
- Gallego, S. (2024). *Predicción de precios del sector inmobiliario en zonas costeras del Atlántico en EE. UU., mediante el uso de técnicas de Machine Learning*.
- García, J. (2023). *Inteligencia artificial y revisión fáctica en la Jurisdicción social: integrando GPT-4 en la resolución de los recursos de suplicación*.
- García José. (2024). *Breve estado del arte en minería de información social*.
- García, M. R., De, T., Garrido, S., Belén, A., Celino, V., Gervás, P., Raquel, G.-N., & Ballesteros, H. (2012). *Desarrollo de una aplicación de ayuda para la redacción de textos simplificados*.
<https://hdl.handle.net/20.500.14352/46155>
- González Calvo, J. M. (1990). *Sobre el concepto de morfema*.
<https://dehesa.unex.es:8443/handle/10662/4503>

- González-Delgado, G., & Navarro-Colorado, B. (2024). *The Simplification of the Language of Public Administration: The Case of Ombudsman Institutions*. 125-133.
- Granados, J. A. T. (1994). *Medidas de prevalencia y relación incidencia-prevalencia*.
- Hael, V. (2016). Comprensión de hiperónimos por parte de sordos a través del método logogenia. *Dialnet.Unirioja.Es*. <https://dialnet.unirioja.es/descarga/articulo/7434122.pdf>
- Herrera Arjonilla, Á. (2024). *TECNOLOGÍAS DEL LENGUAJE PARA TUTORIZACIÓN DOCENTE CON AGENTES CONVERSACIONALES*.
- Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. In *Geoscientific Model Development* (Vol. 15, Issue 14, pp. 5481-5487). Copernicus GmbH. <https://doi.org/10.5194/gmd-15-5481-2022>
- Hueso González, A., & Cascant i Sempere, M. J. (2012). Metodología y técnicas cuantitativas de investigación. *MINISTERIO DE EDUCACION*. <https://repositorio.minedu.gob.pe/handle/20.500.12799/4600>
- Jalbert, A., Neath, I., Bireta, T. J., & Surprenant, A. M. (2011). When does length cause the word length effect? *Journal of Experimental Psychology: Learning Memory and Cognition*, 37(2), 338-353. <https://doi.org/10.1037/A0021804>
- Katz, D. M., Bommarito, M. J., Gao, S., & Arredondo, P. (2024). *GPT-4 passes the bar exam*. <https://doi.org/10.1098/rsta.2023.0254>
- Kaur, J. , & B. P. K. (2018). A systematic review on stopword removal algorithms. . *International Journal on Future Revolution in Computer Science & Communication Engineering*.
- Kelious, A., Constant, M., & Coeur, C. (2024). *Investigating strategies for lexical complexity prediction in a multilingual setting using generative language models and supervised approaches*.
- Keskin, İ., & Topuz, D. (2021). An Application of Fuzzy Pearson Correlation Methods in Animal Sciences. *Selcuk Journal of Agricultural and Food Sciences*, 35(3), 265-271. <https://doi.org/10.15316/sjafs.2021.256>
- Kew, T., Chi, A., Vásquez-Rodríguez, L., Agrawal, S., Aumiller, D., Alva-Manchego, F., & Shardlow, M. (2023). *BLESS: Benchmarking Large Language Models on Sentence Simplification*. <https://chat.openai.com/>
- Khakzad Shahandashti kimya, K., Sivakumar, M., Mahdi Mohajer, M., Belle, A. B., Wang, S., Lethbridge, T. C., Khakzad Shahandashti, K., & Mahdi Mo-hajer, M. (2024). *Evaluating the Effectiveness of GPT-4 Turbo in Creating Defeaters for Assurance Cases*. <https://arxiv.org/abs/2401.17991v1>
- Kher, A., Johnson, S., & Griffith, R. (2017). Readability Assessment of Online Patient Education Material on Congestive Heart Failure.

- Advances in Preventive Medicine*, 2017, 1-8.
<https://doi.org/10.1155/2017/9780317>
- Konkol, M. (2016). *UWB at SemEval-2016 Task 11: Exploring Features for Complex Word Identification*. 1038-1041.
- Larbi, C. C., Djaber, R., & Ismail, H. (2023). *Gunning Fog Index and Automated UML Generation*. <https://doi.org/10.21203/RS.3.RS-3597953/V1>
- Lasecki, W. S., Rello, L., & Bigham, J. P. (2015). Measuring text simplification with the crowd. *W4A 2015 - 12th Web for All Conference*. <https://doi.org/10.1145/2745555.2746658>
- Lewis Tunstall, L. von W. T. W. (2022). *Natural Language Processing with Transformers, Revised Edition* (O'Reilly Media, Ed.).
- Lewis-Beck, M. S., & Skalaban, A. (1990). The R-Squared: Some Straight Talk. *Cambridge University Press*, 2(Political Analysis), 153-171.
- Lin, Z., Du, Y., D'mello, S. K., Luttges, B. L., Shi, Z., & Wang, H. (2024). *Beyond principlism: Practical strategies for ethical AI use in research practices*. <https://doi.org/10.1007/s43681-024-00585-5>
- Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., Wu, Z., Zhao, L., Zhu, D., Li, X., Qiang, N., Shen, D., Liu, T., & Ge, B. (2023). *Summary of ChatGPT-Related Research and Perspective Towards the Future of Large Language Models*. <https://doi.org/10.1016/j.metrad.2023.100017>
- Llisterri, J. (1989). Los sistemas vocálicos: tipología, universales y explicación fonética. *Anuario Del Seminario de Filología Vasca "Julio de Urquijo,"* 23(2), 435-446.
<https://doi.org/10.1387/ASJU.8055>
- LOTAIP. (2004). *LEY ORGANICA DE TRANSPARENCIA Y ACCESO A LA INFORMACION PUBLICA*. 2004.
<https://www.educacionsuperior.gob.ec/wp-content/uploads/downloads/2014/09/LOTAIP.pdf>
- Lotman, I. M., De, U., Traducido, T., Darío, R., & Arcila, F. (2002). *El símbolo en el sistema de la cultura*.
- Lúcar Gonzalez, B., & Fernando, L. (2019). *Frecuencia relativa y acumulada Trabajo de Suficiencia Profesional Para optar el Título Profesional de*.
- Lyons III, P. J. (2018). *The Impact of Lexical Complexity on the Public's Understanding The Impact of Lexical Complexity on the Public's Understanding of Governmental Communication of Governmental Communication*.
https://digitalcommons.buffalostate.edu/mpa_projects/28
- Maddela, M., & Xu, W. (2018). *A Word-Complexity Lexicon and A Neural Readability Ranking Model for Lexical Simplification*. 3749-3760.
<https://github.com/>
- Mao, R., Chen, G., Zhang, X., Guerin, F., & Cambria, E. (2023). *GPTEval: A Survey on Assessments of ChatGPT and GPT-4*. Retrieved

- December 21, 2024, from <https://docs.cohere.com/docs/command-beta>
- McCarthy, P. M., & Jarvis, S. (2010). MTL-D, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42(2), 381-392. <https://doi.org/10.3758/BRM.42.2.381/METRICS>
- Meneses Pacheco, C. (2018). Significado De La Fe Pública En La Prueba Por Medio De Documentos Públicos. *Revista de Derecho (Coquimbo)*, 25(1), 181-230. <https://doi.org/10.4067/S0718-97532018000100181>
- Mo, K., & Hu, R. (2024). *ExpertEase: A Multi-Agent Framework for Grade-Specific Document Simplification with Large Language Models*.
- Mosquera, A. (2021). *Alejandro Mosquera at SemEval-2021 Task 1: Exploring Sentence and Word Features for Lexical Complexity Prediction*. 554-559. Retrieved December 21, 2024, from <https://github.com/sheffieldnlp/cwi>
- North, K., Zampieri, M., & Shardlow, M. (2023). Lexical Complexity Prediction: An Overview. *ACM Computing Surveys*, 55(9). <https://doi.org/10.1145/3557885/ASSET/F0C06F93-9FD9-4E29-AA22-C73EFC8CF458/ASSETS/GRAPHIC/CSUR-2021-0414-INLINE2.JPG>
- Ochoa Sierra, L., & Cueva Lobelle, A. (2020). COMPLEJIDAD LINGÜÍSTICA: REVISIÓN BIBLIOGRÁFICA. *Signo y Pensamiento*, 39(77). <https://doi.org/10.11144/JAVERIANA.SYP39-77.CLRB>
- Ono, D., Dickson, D. W., & Koga, S. (2024). Evaluating the efficacy of few-shot learning for GPT-4Vision in neurodegenerative disease histopathology: A comparative analysis with convolutional neural network model. *Neuropathology and Applied Neurobiology*, 50(4). <https://doi.org/10.1111/nan.12997>
- Ortega-Mendoza, R. M., Villaseñor-Pineda, L., Montes-Y-Gómez, M., Amaro, K. R., Ortega González, E. V., Hiram, F., & Castro, C. (2007). *Construcción Automática de Catálogos de Hipónimos a partir de Texto no Estructurado*.
- OpenAI. (2023). *GPT-4 Technical Report*.
- OpenAI, :, Hurst, A., Lerer, A., Goucher, A. P., Perelman, A., Ramesh, A., Clark, A., Ostrow, A., Welihinda, A., Hayes, A., Radford, A., Ądry, A., Baker-Whitcomb, A., Beutel, A., Borzunov, A., Carney, A., Chow, A., Kirillov, A., ... Malkov, Y. (2024). *GPT-4o System Card*. <https://arxiv.org/abs/2410.21276v1>
- Ortiz Zambrano, J., MontejoRáez, A., Lino Castillo, K. N., Gonzalez Mendoza, O. R., & Cañizales Perdomo, B. C. (2020). VYTEDU-CW: Difficult Words as a Barrier in the Reading Comprehension of University Students. *Advances in Intelligent Systems and Computing*, 1066, 167-176. https://doi.org/10.1007/978-3-030-32022-5_16

- Ortiz-Zambrano, J. A., Espín-Riofrío, C. H., & Montejo-Ráez, A. (2024). Deep Encodings vs. Linguistic Features in Lexical Complexity Prediction. *Neural Computing and Applications*. <https://doi.org/10.1007/s00521-024-10662-9>
- Ortiz-Zambrano, J. A., Espin-Riofrío, C., & Montejo-Ráez, A. (2023). *LegalEc: A New Corpus for Complex Word Identification Research in Law Studies in Ecuatorian Spanish*. <https://doi.org/10.26342/2023-71-19>
- Ortiz-Zambrano, J. A., Espin-Riofrío, C., & Montejo-Ráez, A. (2023). LegalEc: A New Corpus for Complex Word Identification Research in Law Studies in Ecuatorian Spanish. *Procesamiento Del Lenguaje Natural*, 71, 247-259. <https://doi.org/10.26342/2023-71-19>
- Ortiz-Zambrano, J. A., & Montejo-Ráez, A. (2020). Overview of ALexS 2020: First Workshop on Lexical Analysis at SEPLN. <https://www.ujaen.es/>
- Ortiz-Zambrano, J., Espin-Riofrío, C., & Montejo-Ráez, A. (2022). Transformers for Lexical Complexity Prediction in Spanish Language. *Procesamiento Del Lenguaje Natural*, 69, 177-188. <https://doi.org/10.26342/2022-69-15>
- Ortiz-Zambrano, J., Espin-Riofrío, C., & Montejo-Ráez, A. (2023). SINAI Participation in SimpleText Task 2 at CLEF 2023: GPT-3 in Lexical Complexity Prediction for General Audience Notebook for the SimpleText Lab at CLEF 2023. <http://ceur-ws.org>
- Ortiz-Zambrano, J., Espin-Riofrío, C., & Montejo-Ráez, A. (2024a). SINAI Participation in SimpleText Task 2 at CLEF 2024: Zero-shot Prompting on GPT-4-Turbo for Lexical Complexity Prediction Notebook for the SimpleText Lab at CLEF 2024. <https://openai.com/>
- Ortiz-Zambrano, J., Espin-Riofrío, C., & Montejo-Ráez, A. (2024b). SINAI Participation in SimpleText Task 2 at CLEF 2024: Zero-shot Prompting on GPT-4-Turbo for Lexical Complexity Prediction Notebook for the SimpleText Lab at CLEF 2024. <https://openai.com/>
- Ortiz-Zambrano, J., & Montejo-Ráez, A. (2021). SINAI at SemEval-2021 Task 1: Complex word identification using Word-level features. 126-129. Retrieved December 20, 2024, from <https://pypi>.
- Ortiz-Zambrano, J., & Montejo-Ráez, A. (2021). SINAI at SemEval-2021 Task 1: Complex word identification using Word-level features. <https://pypi>.
- Petrov, S., Das, D., & Mcdonald, R. (2011). arXiv:1104.2086v1 [cs.CL] 11 Apr 2011. <http://www.sejong.or.kr>
- Poblete, C. A., & González, P. F. (2018). Una mirada a l'ús de llenguatge clar en l'àmbit judicial llatinoamericà. *Revista de Llengua i Dret*, 69, 119-138. <https://doi.org/10.2436/RLD.169.2018.3051>
- Phoocharoensil, S. (2010). A Corpus-Based Study of English Synonyms. *International Journal of Arts and Sciences*, 3(10), 227-245.

- Pottier, B. (1961). Sobre el concepto de verbo auxiliar. *Nueva Revista de Filología Hispánica*.
- Ramalho, A. / A. :, Filipa, M., Barberá, H., María Cotutor, S., Domingo, M., & Laura, M. (2024). *Aplicación de gestión y análisis de documentos de contratación pública en las fases de diseño y construcción mediante el uso de técnicas de IA*.
- Ramón, J., & Corredera, C. (2023). *Inteligencia artificial generativa*.
- Ruohonen, J. (2021). *Assessing the Readability of Policy Documents on the Digital Single Market of the European Union*. <https://doi.org/10.1109/ICEDEG52154.2021.9530996>
- Schober, P., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia and Analgesia*, 126(5), 1763-1768. <https://doi.org/10.1213/ANE.0000000000002864>
- Shong, N., & Bs, C. (2008). PEARSON'S VERSUS SPEARMAN'S AND KENDALL'S CORRELATION COEFFICIENTS FOR CONTINUOUS DATA.
- Sims-Williams, H. (2022). *Token frequency as a determinant of morphological*. <https://doi.org/10.1017/S0022226721000438>
- Stamatatos, E. (2009). *Intrinsic Plagiarism Detection Using Character n-gram Profiles*.
- Strauss, U., Grzybek, P., & Altmann, G. (2007). Word Length and Word Frequency. *Contributions to the Science of Text and Language*, 277-294. https://doi.org/10.1007/978-1-4020-4068-9_13
- Tatachar, A. V. (2021). Comparative Assessment of Regression Models Based On Model Evaluation Metrics. *International Research Journal of Engineering and Technology*. www.irjet.net
- Thorndike, E. (1921). *The teacher's word book*. New York Teachers College, Columbia University.
- Tinker, M. (1963). *Legibility of Print* (Iowa State University Press, Ed.).
- Trott, S., & Rivière, P. D. (2024). *Measuring and Modifying the Readability of English Texts with GPT-4*. Retrieved December 28, 2024, from <https://osf.io/6hmej>.
- Trujillo, I. T., Manuel, J., & Troncoso, C. (2024). *Evaluación Comparativa de las Arquitecturas Mamba y Transformers*.
- Torner Castells, S., & Dra Paz Battaner Dra Louise McNally, D. M. (2005). *Aspectos de la semántica de los adverbios de modo en español*.
- Tsaprouni, E., & Manouilidou, C. (2025). The Role of Aspect During Deverbal Word Processing in Greek. *Journal of Psycholinguistic Research*, 54(1), 7. <https://doi.org/10.1007/S10936-024-10112-6>
- Van Guido. (2024). *Chat GPT 4 presentado: Una Guía Integral para Principiantes para Dominar Conversaciones con IA* (D. O'Brien, Ed.).

- Vargas Cordero, Z. R. ,. (2009). *LA INVESTIGACIÓN APLICADA: UNA FORMA DE CONOCER LAS REALIDADES CON EVIDENCIA CIENTÍFICA*.
- Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention Is All You Need*.
- Wani, N., Mathias, S., Aanand Gajjam, J., & Bhattacharyya, P. (2018). *The Whole is Greater than the Sum of its Parts: Towards the Effectiveness of Voting Ensemble Classifiers for Complex Word Identification*. 200-205.
- White, R. H. (2014). *Lexical richness in adolescent writing, insights from the classroom: An L1 vocabulary development study*. <https://doi.org/10.26686/WGTN.17007397.V1>
- Xu, G., Hammani, K., Chabchoub, A., Dudley, J. M., Kibler, B., & Finot, C. (2019). *Phase evolution of Peregrine-like breathers in optics and hydrodynamics*.
- Zhou, S., Jeong, H., & Green, P. A. (2017). How consistent are the best-known readability equations in estimating the readability of design standards? *IEEE Transactions on Professional Communication*, 60(1), 97-111. <https://doi.org/10.1109/TPC.2016.2635720>

ÍNDICE GENERAL

Requisitos para el desarrollo del proyecto.....	109
Herramientas utilizadas para el proyecto de desarrollo ...	109
Visual Studio Code.....	109
Python.....	109
Project-GPT-4-few-shot	110
Estructura de las carpetas.....	110
Carpeta corpus.....	111
Carpeta prompt_examples.....	111
Carpeta proyect_modules	112
main.py	112
GPT-4.py.....	113
requeriments.txt	119
Project-GPT-4-zero-shot	119

ÍNDICE DE FIGURAS

Fig. 1 Visual Studio Code	109
Fig. 2 Python	110
Fig. 3 Estructura de las carpetas.....	111
Fig. 4 Carpeta corpus	111
Fig. 5 Carpeta prompt_examples	111
Fig. 6 Carpeta proyect_modules.....	112
Fig. 7 main.py	113
Fig. 8 GPT-4.py	113
Fig. 9 Funciones __prompt_format e __imprimir_fila.....	114
Fig. 10 Funciones __imprimir_fila_porcent e __asig_etiqueta	115
Fig. 11 Funciones __start_1 y __start_2	116
Fig. 12 Función __filtro	116
Fig. 13 Función _seleccionar_palabras_complejas	117
Fig. 14 Función __evaluar	117
Fig. 15 Función data_to_process	118
Fig. 16 Función process_all	118
Fig. 17 requeriments.txt	119
Fig. 18 Project-GPT-4-zero-shot	119

El presente manual técnico se explican los códigos correspondientes a las ejecuciones de los modelos GPT-4 Turbo y GPT-4o, se describirá cada proyecto dependiendo de su técnica zero-shot learning y few-shot learning.

Requisitos para el desarrollo del proyecto

Los requisitos mínimos de hardware que se recomienda tener en el equipo de escritorio PC o laptop son los siguientes componentes.

- Core i5 - 10100T CPU
- 6GB de RAM
- Windows 10 o 11 de 64 bits

Herramientas utilizadas para el proyecto de desarrollo

Se utilizaron las siguientes herramientas para las ejecuciones de los modelos.

- Visual Studio Code
- Python

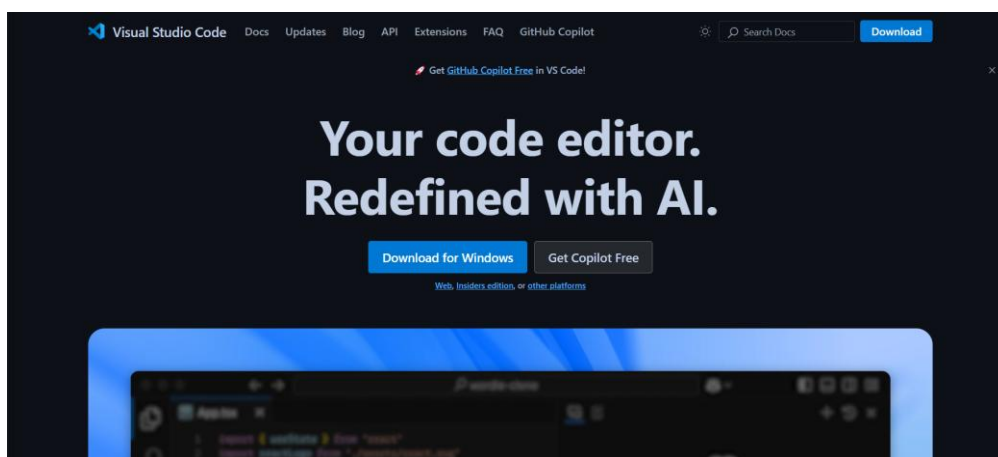
Visual Studio Code

VSC es un editor de código fuente gratuito y de código abierto desarrollado por Microsoft, es ampliamente utilizado por los desarrolladores de software en diversos sistemas operativos, como Windows, macOS y Linux. El siguiente enlace es para su descarga: <https://code.visualstudio.com/>

Fig.

1

Visual Studio Code



Nota: Elaborado por los autores

Python

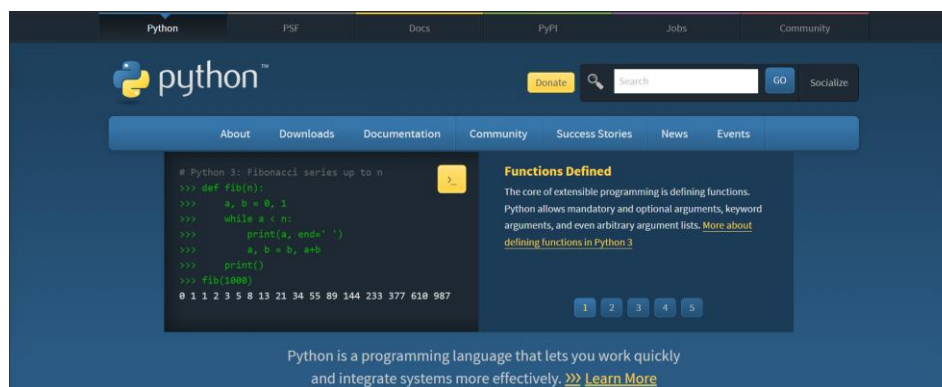
Es un lenguaje de programación de alto nivel, interpretado y de propósito general, conocido por su sintaxis sencilla que facilita la escritura de código. Es multiplataforma, es decir, se puede instalar

en sistemas operativos como Windows, macOS y Linux. Se recomienda la descarga de la versión 3.12.0 para evitar problemas con las bibliotecas: <https://www.python.org/>

Fig.

2

Python



Nota: Elaborado por los autores

Project-GPT-4-few-shot

Empezare con la explicación del código del proyecto GPT-4 que emplea la técnica few-shot learning para las ejecuciones de los sucesores GPT-4 Turbo y GPT-4o.

Estructura de las carpetas

A continuación, se describirá cada una de las carpetas existentes en el proyecto:

corpus: Carpeta que contiene los datos para el proyecto.

prompt_examples: Carpeta que contiene el prompt.

project_modules: Carpeta que contiene los módulos del proyecto.

resultados: Carpeta que contiene los resultados del proyecto.

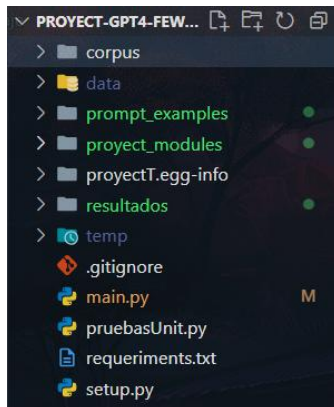
.gitignore: Archivo que indica los archivos que no se deben subir al repositorio de Git.

main.py: Archivo principal del proyecto.

requeriments.txt: Archivo que lista las dependencias del proyecto.

Fig.
Estructura de las carpetas

3



Nota: Elaborado por los autores

Carpeta corpus

La carpeta corpus contiene el corpus de datos que se va a pasar a los modelos para las ejecuciones, las columnas pasadas son id, corpus, sentence, token y complexity.

Fig.
Carpeta corpus

4

	A	B	C	D
1	source	id	corpus	sentence
2		1	6,075	Municipio - Tramites - TEXTO 0060 TRAMITES EN LA BIBLIOTECA MUNICIPAL.txt
3		2	6,076	Municipio - Tramites - TEXTO 0060 TRAMITES EN LA BIBLIOTECA MUNICIPAL.txt
4		3	6,077	Municipio - Tramites - TEXTO 0060 TRAMITES EN LA BIBLIOTECA MUNICIPAL.txt
5		4	6,093	Municipio - Tramites - TEXTO 0062 TRAMITES EN LA BIBLIOTECA MUNICIPAL.txt

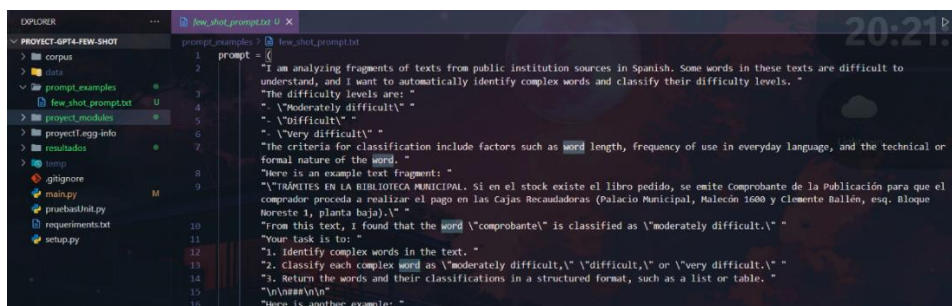
Nota: Elaborado por los autores

Carpeta prompt_examples

Comencemos con la carpeta prompt_examples, esta contiene un .txt con ejemplos para el uso de la técnica de few-shot, el archivo few_shot_prompt.txt contiene 20 ejemplos que ayudaran a los modelos identificar las palabras complejas.

Fig.
Carpeta prompt_examples

5



Nota: Elaborado por los autores

Carpeta `project_modules`

Ahora nos vamos a centrar en la carpeta `project_modules`, esta contiene archivos importantes, el primero de ellos es la carpeta `__pycache__`, la cual, se crea por defecto por Python al crear el proyecto para almacenar el bytecode compilado para una ejecución más rápida.

`__init__.py`: Contiene información sobre el título del proyecto, la versión y una descripción.

`file_metrics.py`: Contiene la función para calcular las métricas, genera un `.xlsx` con los resultados finales de todas las ejecuciones.

`file_operation.py`: Contiene varias funciones que gestionan datos temporales.

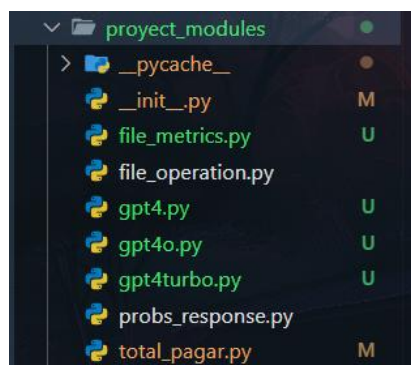
`gpt4.py`, `gpt4o.py` y `gpt4turbo.py`: Es el código que se utilizan para el procesamiento del corpus, en este caso las ejecuciones se centraron en los tres modelos.

`total_pagar.py`: Contiene una función que calcula el costo total de procesamiento de un conjunto de datos basándose en la cantidad de tokens en los prompts y la longitud de las respuestas generadas.

Fig.

6

Carpeta `project_modules`



Nota: Elaborado por los autores

main.py

En este archivo principal, en el main se carga el prompt, se llama al modelo y se realiza la ejecución, las variables `minimo` y `maximo` definen el rango de datos a analizar, luego se establecen las rutas de los archivos de datos de entrenamiento y prueba. En resumen, se crea un objeto dependiendo del modelo (`Gpt4`, `Gpt4o` y `Gpt4Turbo`) que se inicializa con los datos, el prompt, la API key y se establece si se deben cargar datos previamente, luego se

utiliza el método `process_all` del objeto modelo para realizar el análisis de los datos y generar predicciones, estos resultados del análisis se guardan.

Fig.
main.py

7

```

13
14 def cargar_prompt(archivo):
15     with open(archivo, 'r', encoding='utf-8') as file:
16         return file.read()
17
18
19 if __name__ == "__main__":
20     parser = argparse.ArgumentParser()
21     parser.add_argument("-p", "--percent", help="Muestra porcentaje de las respuestas", action="store_true")
22     parser.add_argument("-l", "--load", help="Carga datos si existio algun error en el codigo", action="store_true")
23     args = parser.parse_args()
24
25     minimo = 0
26     maximo = 7812
27     file_corpus_train = r"corpus\Adminlex_single_train_normalizacion.xlsx"
28     file_corpus_test = r"corpus\Adminlex_single_train_normalizacion.xlsx"
29
30     archivo_prompt = r"prompt_examples\zero_shot_prompt.txt"
31     prompt = cargar_prompt(archivo_prompt)
32
33     df = pd.read_excel(file_corpus_test, engine='openpyxl')
34     df.loc[df["token"].isnull(), "token"] = "null"
35     datos = df.loc[minimo:maximo, ["id", "source", "sentence", "token", "complexity", "escala"]]
36     # key = "sk-c04an2xjF17PfhfLUUNT3B1bkF7tGK70EOCuv9ukkhfLMHt"
37     key = "sk-proj-2HY_UNTU1q_4FwObucdt4JQokKSZzgcP3buiRwWxfbly3o1dyU-FNcLpik82U39f8ebGeITqT3B1bkFJ2u5yagjKdrii7RdhyqeQN3Pp1kk-"
38     # gpt = Gpt4(datos, prompt, key, load=args.load)
39     # gpt = Gpt4turbo(datos, prompt, key, load=args.load)
40     gpt = Gpt4o(datos, prompt, key, load=args.load)
41     # calcular_total_pagar(prompt, datos, cost=0.03)
42     gpt.process_all(file_path=file_corpus_train, save_result=True, percent=args.percent)

```

Nota: Elaborado por los autores

GPT-4.py

En este código se define una clase llamada 'Gpt4' que representa el modelo a analizar y presentar datos relacionados con las diferentes estrategias: STRAT_1, STRAT_2, el Constructor `__init__` toma los argumentos `data`, `prompt`, `key`, y `load`. Se inicializa las variables de clase `_plantilla_resultados`, `_plantilla_porcentaje`, y `_medias`, y el diccionario `_rango_escalas` se define la categoría de los niveles de dificultad.

Fig.
GPT-4.py

8

```

class Gpt4:
    STRAT_1 = 'strat1'
    STRAT_2 = 'strat2'

    def __init__(self, datos, prompt, key, load=False):
        self._plantilla_resultados = "{:^5} {:^20} {:^20} {:^30} {:^30} {:^30} {:^30} {:^30}"
        self._plantilla_porcentaje = "{:^5} {:^20} {:^20} {:^20}"
        self._medias = {}
        self._datos = datos
        self._prompt = prompt
        self._key = key
        self.load = load
        self._rango_escalas = {
            'moderately difficult': (0, 0.3333),
            'difficult': (0.3334, 0.6666),
            'very difficult': (0.6667, 1)
        }

```

Nota: Elaborado por los autores

Funciones `__prompt_format` e `__imprimir_fila`:

`__prompt_format`: Retorna el prompt formateado, la función formatea el prompt (`few_shot_prompt.txt`) reemplazando los marcadores de posición con los valores: `@recurso => source`, `@oracion => sentence` y `@aEvaluar => token`.

`__imprimir_fila`: Utiliza el método llamado `format` para insertar datos en una plantilla de resultados, es decir, esta función imprime una fila de datos.

Fig.

9

Funciones `__prompt_format` e `__imprimir_fila`

```
def __prompt_format(self, source, sentence, token):
    prompt = self.__prompt
    prompt = prompt.replace("@recurso", f"\{source}\")
    prompt = prompt.replace("@oracion", f"\{sentence}\")
    prompt = prompt.replace("@aEvaluar", f"\{token}\")
    return prompt

def __imprimir_fila(
    self, indice, respuesta_gpt4, rango, complejidad_gpt4,
    complejidad, complejidad_escalas, comparacion, palabras_complejas):
    token = self.__datos["token"][indice]

    print(self.__plantilla_resultados.format(
        indice, token, respuesta_gpt4, rango,
        complejidad_gpt4, complejidad, complejidad_escalas,
        comparacion, palabras_complejas
    ))
```

Nota: Elaborado por los autores

Funciones `__imprimir_fila_porcent` e `__asig_etiqueta`:

`__imprimir_fila_porcent`: Imprime una fila de datos utilizando una plantilla, con los siguientes parámetros: `self`, `índice`, `respuesta_gpt4`, `respuesta_complex`, `opciones` y `Acciones`, utilizando el método `format` de la plantilla `__plantilla_porcentaje` para insertar los datos en la plantilla.

`__asig_etiqueta`: Asigna una etiqueta de dificultad a un valor, se define una variable `escala` para almacenar la etiqueta de dificultad y comparar ese valor con los rangos de dificultad.

Fig.**10**Funciones `__imprimir_fila_porcent` e `__asig_etiqueta`

```

def __imprimir_fila_porcent(self, indice, respuesta_gpt4, respuesta_complex, opciones):
    token = self.__datos["token"][indice]

    print(self.__plantilla_porcentaje.format(
        indice, token, respuesta_gpt4, respuesta_complex, opciones[0],
        opciones[1], opciones[2], opciones[3]))

def __asig_etiqueta(self, valor):
    escala = ""
    if valor <= 0.3333:
        escala = "moderately difficult"
    elif 0.3334 < valor <= 0.6666:
        escala = "difficult"
    elif 0.6667 < valor <= 1:
        escala = "very difficult"
    return escala

```

Nota: Elaborado por los autoresFunciones `__start__1` y `__start__2`:

`__start__1`: Utiliza la función `reduce` que toma dos argumentos: `lambda` que define la operación a realizar en cada iteración y el rango de escalas a iterar, luego `lambda` suma los valores de la iteración actual y la anterior y divide por 2.

`__start__2`: Esta función procesa un `.xlsx`, para cada valor de la lista de claves, selecciona las filas del `DataFrame` donde la columna 'escala' es igual al valor actual, se calcula el promedio de la columna deseada y lo almacena en el diccionario bajo la clave correspondiente.

Fig.**11**Funciones `__start__1` y `__start__2`

```

def __strat_1(self, valor_escalas):
    return reduce(lambda x, y: (x + y) / 2, self.__rango_escalas.get(valor_escalas))

def __strat_2(self, name_file):
    diccionario = {}

    if not os.path.exists("data"):
        os.mkdir("data")

    try:
        tf = open("data/promedio.json", "r")
        diccionario = json.load(tf)
        no_file = False
    except FileNotFoundError:
        no_file = True

    if no_file:
        df = pd.read_excel(name_file)
        for valor in list(self.__rango_escalas.keys()):
            aux = df[df["escala"] == valor]
            calculo = aux["complexity"].mean()
            diccionario[valor] = calculo
        tf = open("data/promedio.json", "w")
        json.dump(diccionario, tf)
        tf.close()

    self.__means = diccionario

```

Nota: Elaborado por los autores

Función `__filtro`: Toma dos argumentos 'self' y 'respuesta_gpt4', esta función recorre las claves del diccionario '`__rango_escalas`' y verifica dos condiciones: la primera es si el número de palabras en la clave es 2 y la cantidad de veces que aparece la clave en la cadena `respuesta_gpt4` es mayor o igual a 1, la función devuelve la clave, la segunda es si la cantidad de veces que aparece la clave en la cadena `respuesta_gpt4` es mayor o igual a 1, la función asigna la clave a la variable resultado y luego la devuelve.

Fig.**12**Función `__filtro`

```

def __filtro(self, respuesta_gpt4):
    resultado = ""

    for valor_escalas in list(self.__rango_escalas.keys()):
        n_palabras = len(valor_escalas.split())
        if n_palabras == 2 and respuesta_gpt4.count(valor_escalas) >= 1:
            return valor_escalas
        elif respuesta_gpt4.count(valor_escalas) >= 1:
            resultado = valor_escalas

    return resultado

```

Nota: Elaborado por los autores

Función `__seleccionar_palabras_complejas`: Esta función utiliza la API de OpenAI para identificar palabras complejas en una oración dada y devuelve una lista de esas palabras. Para la

identificación de palabras complejas se utiliza el modelo gpt-3.5-turbo-instruct porque es menos costoso, y tiene un buen rendimiento en su tarea de identificación de palabras complejas.

Fig. 13
Función `_seleccionar_palabras_complejas`

```
def _seleccionar_palabras_complejas(self, sentence, top_n=5):
    try:
        client = OpenAI(api_key=self.__key)
        response = client.completions.create(
            model="gpt-3.5-turbo-instruct",
            prompt=f"Write up to {top_n} the most complex words in the following sentence: \"{sentence}\"",
            temperature=0.5,
            max_tokens=50,
            top_p=1,
            frequency_penalty=0,
            presence_penalty=0,
            logprobs=5,
            stop=["\n"]
        )

        palabras_complejas = response.choices[0].text.strip().split(",")
        print(sentence)
        print(palabras_complejas)
        return palabras_complejas
    except Exception as e:
        print(f"Error al identificar palabras complejas: {str(e)}")
        return []
```

Nota: Elaborado por los autores

Función `__evaluar`: Esta función recibe dos argumentos, una referencia al objeto de la función 'self' y la instrucción a evaluar 'orden'. En conclusión, se envía una solicitud de la API de OpenAI para que el modelo genere una respuesta en base a la instrucción dada. Es decir, evalúa la palabra compleja.

Fig. 14
Función `__evaluar`

```
def __evaluar(self, orden):
    client = OpenAI(
        api_key=self.__key,
    )

    response = client.completions.create(
        model="gpt-4",
        prompt=orden,
        temperature=0.2,
        max_tokens=5,
        top_p=1,
        frequency_penalty=0,
        presence_penalty=0,
        logprobs=5,
        # top_logprobs=5,
        stop=["\n"]
    )

    respuesta = response.choices[0].text
    prob_tokens = response.choices[0].logprobs.top_logprobs
    return respuesta, prob_tokens
```

Nota: Elaborado por los autores

Función `data to process`: Esta función procesa los datos para su posterior análisis, se compone de varios bloques de código que ejecutan diferentes acciones dependiendo de la condición del

argumento `load_da`, como la respuesta del modelo, el rango de la complejidad, la complejidad que el modelo asigno a la palabra identificada como compleja y la comparación.

Fig.

15

Función `data_to_process`

```
def data_to_process(self):
    load_da, minimo, maximo = load_data_temp() if self.load else (None, None, None)

    if load_da is None:
        to_process = self.__datos
        to_process["Respuesta GPT-4_zero_shot"] = None
        to_process["Rango GPT-4_zero_shot"] = None
        to_process["Complejidad GPT-4_zero_shot"] = 0.0
        to_process["comparacion"] = None
        for i in range(5):
            to_process[f"Porcentaje {i + 1}"] = ""
        minimo = 0
        maximo = self.__datos.shape[0] - 1

    elif load_da is not None:
        to_process = load_da
        to_process = pd.concat([to_process, self.__datos[minimo:]], ignore_index=True)
    else:
        sys.exit("Error de ingreso de parametros")

    return to_process, minimo, maximo
```

Nota: Elaborado por los autores

Función `process_all`: Esta función toma el corpus como entrada, determina el rango de datos a procesar y luego calcula la complejidad, después se comparan los resultados obtenidos del modelo y el valor real, y se calculan las métricas estadísticas y las correlaciones. Al final se definen las columnas del Dataframe y se crea un .xlsx con el nombre del resultado.

Fig.

16

Función `process_all`

```
def process_all(self, file_path="", version=False, save_result=False, percent=False):
    if file_path != "":
        self.__strat_2(file_path)

    resultado, minimo, maximo = self.data_to_process()
```

Nota: Elaborado por los autores

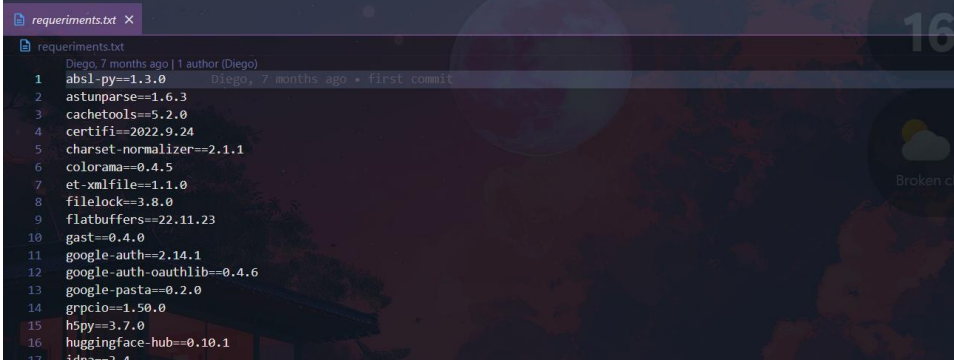
requeriments.txt

Este archivo contiene la lista de dependencias utilizadas en el proyecto y las versiones necesarias para que el proyecto funcione correctamente.

Fig.

17

requeriments.txt



```

requirements.txt
Diego, 7 months ago | 1 author (Diego)
1  absl-py==1.3.0
2  astunparse==1.6.3
3  cachetools==5.2.0
4  certifi==2022.9.24
5  charset-normalizer==2.1.1
6  colorama==0.4.5
7  et-xmlfile==1.1.0
8  filelock==3.8.0
9  flatbuffers==22.11.23
10 gast==0.4.0
11 google-auth==2.14.1
12 google-auth-oauthlib==0.4.6
13 google-pasta==0.2.0
14 grpcio==1.50.0
15 h5py==3.7.0
16 huggingface-hub==0.10.1

```

Nota: Elaborado por los autores

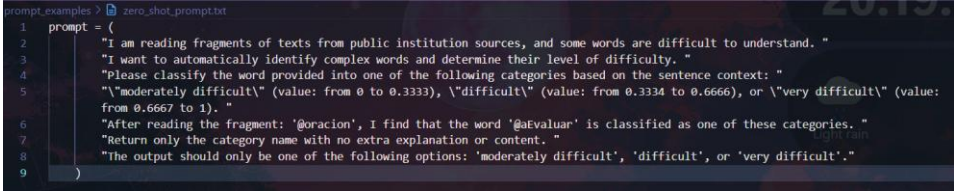
Project-GPT-4-zero-shot

El código para las ejecuciones de los modelos con la técnica de zero-shot learning es similar a la de few-shot, la única diferencia significativa es el prompt, para zero-shot no se proporcionan ejemplos, se le da una instrucción lo mas detallada posible al modelo para que realice la predicción. A continuación, se muestra el prompt de zero-shot.

Fig.

18

Project-GPT-4-zero-shot



```

prompt_examples > zero_shot_prompt.txt
1  prompt = (
2      "I am reading fragments of texts from public institution sources, and some words are difficult to understand. "
3      "I want to automatically identify complex words and determine their level of difficulty. "
4      "Please classify the word provided into one of the following categories based on the sentence context: "
5      "\"moderately difficult\" (value: from 0 to 0.3333), \"difficult\" (value: from 0.3334 to 0.6666), or \"very difficult\" (value: "
6      "from 0.6667 to 1). "
7      "After reading the fragment: '@oracion', I find that the word '@a evaluar' is classified as one of these categories. "
8      "Return only the category name with no extra explanation or content. "
9      "The output should only be one of the following options: 'moderately difficult', 'difficult', or 'very difficult'."

```

Nota: Elaborado por los autores

ÍNDICE GENERAL

Requisitos para la ejecución del proyecto.....	121
Herramientas utilizadas la ejecución del proyecto	121
Visual Studio Code.....	121
Python.....	121
Ejecución del proyecto GPT-4-Zero-Shot.	122
Extraer el .zip del proyecto	122
Abrir el proyecto	122
Abrir cmd	122
Vista del proyecto en vsc	123
Consideraciones	124
Ejecutar proyecto	124

ÍNDICE DE FIGURAS

Fig. 1 Visual Studio Code	121
Fig. 2 Python	122
Fig. 3 Extraer el .zip.....	122
Fig. 4 Abrir símbolo del sistema desde la ruta del proyecto..	123
Fig. 5 Abrir proyecto en vsc desde cmd	123
Fig. 6 Vista del proyecto en vsc	123
Fig. 7 Abrir terminal del proyecto	124
Fig. 8 Ejecutar el proyecto	124

En el presente manual de usuario se explican los pasos para la ejecución de los proyectos GPT-4-Zero-Shot y GPT-4-Few-Shot. A continuación, se explicará la ejecución del proyecto GPT-4-Zero-Shot.

Requisitos para la ejecución del proyecto

Los requisitos mínimos de hardware que se recomienda tener en el equipo de escritorio PC o laptop son los siguientes componentes.

- Core i5 - 10100T CPU
- 6GB de RAM
- Windows 10 o 11 de 64 bits
- Correo electrónico de Gmail

Herramientas utilizadas la ejecución del proyecto

Se utilizaron las siguientes herramientas para las ejecuciones de los modelos.

- Visual Studio Code
- Python

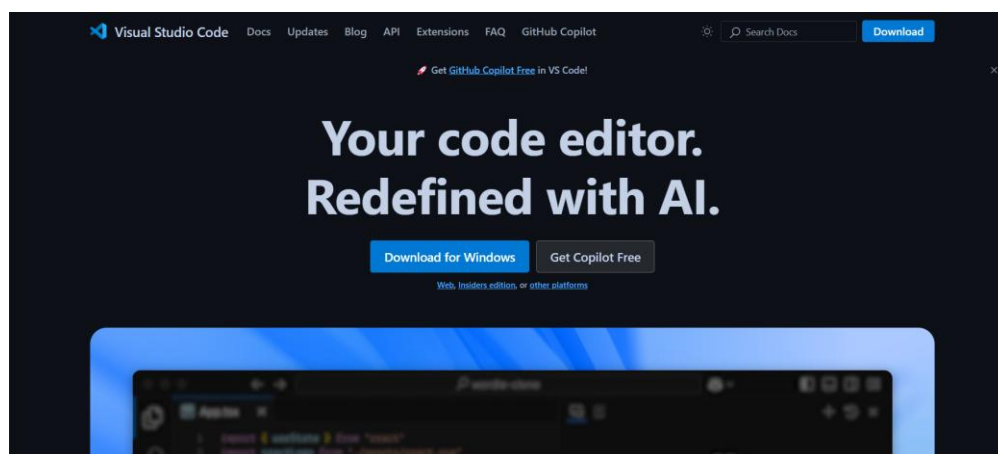
Visual Studio Code

VSC es un editor de código fuente gratuito y de código abierto desarrollado por Microsoft, es ampliamente utilizado por los desarrolladores de software en diversos sistemas operativos, como Windows, macOS y Linux. El siguiente enlace es para su descarga: <https://code.visualstudio.com/>

Fig.

1

Visual Studio Code



Nota: Elaborado por los autores

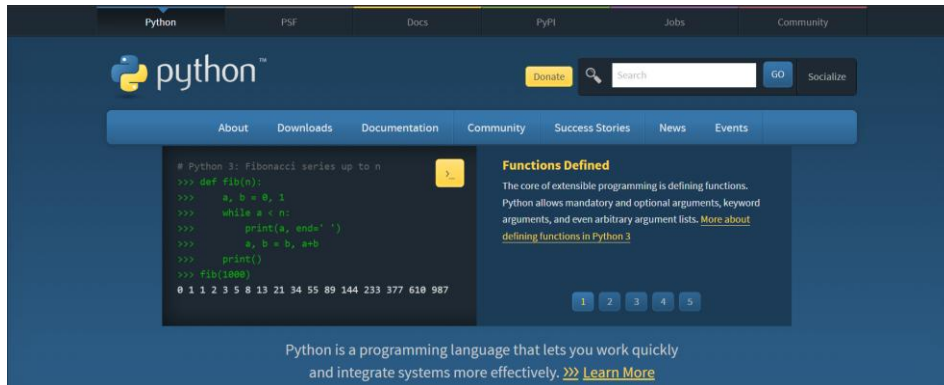
Python

Es un lenguaje de programación de alto nivel, interpretado y de propósito general, conocido por su sintaxis sencilla que facilita la

escritura de código. Es multiplataforma, es decir, se puede instalar en sistemas operativos como Windows, macOS y Linux. Se recomienda la descarga de la versión 3.12.0 para evitar problemas con las bibliotecas: <https://www.python.org/>

Fig.
Python

2



Nota: Elaborado por los autores

Ejecución del proyecto GPT-4-Zero-Shot.

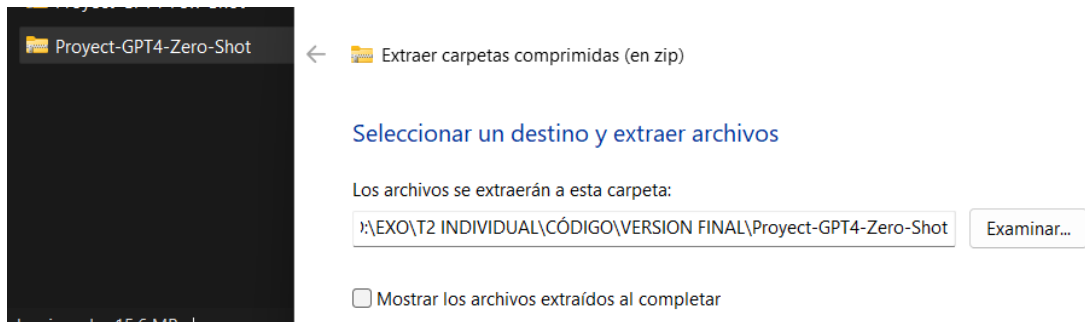
Extraer el .zip del proyecto

Una vez descargado el .zip del proyecto lo extraemos en la ruta deseada para guardar.

Fig.

3

Extraer el .zip



Nota: Elaborado por los autores

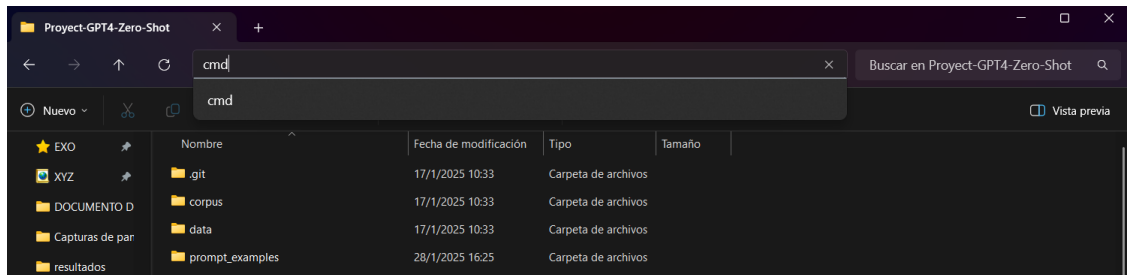
Abrir el proyecto

Una vez que se extrajo el proyecto, lo abrimos en Visual Studio Code, para ello primero nos ubicaremos dentro del contenido del proyecto.

Abrir cmd

Luego escribiremos en la ruta de la carpeta el comando 'cmd'.

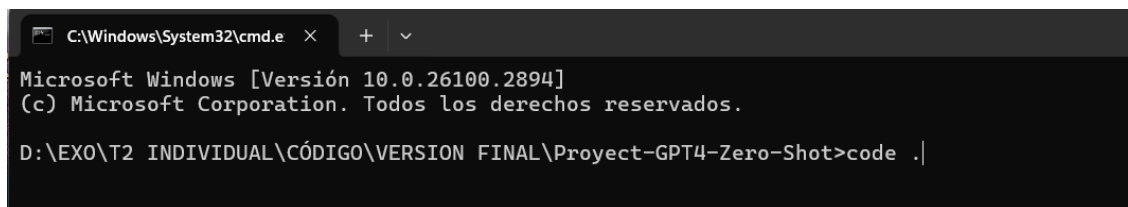
Fig. *Abrir símbolo del sistema desde la ruta del proyecto* **4**



Nota: Elaborado por los autores

Abrir proyecto desde el símbolo del sistema
Dentro del símbolo de sistema (cmd) escribimos el comando 'code .' esto nos abrirá el proyecto en Visual Studio Code.

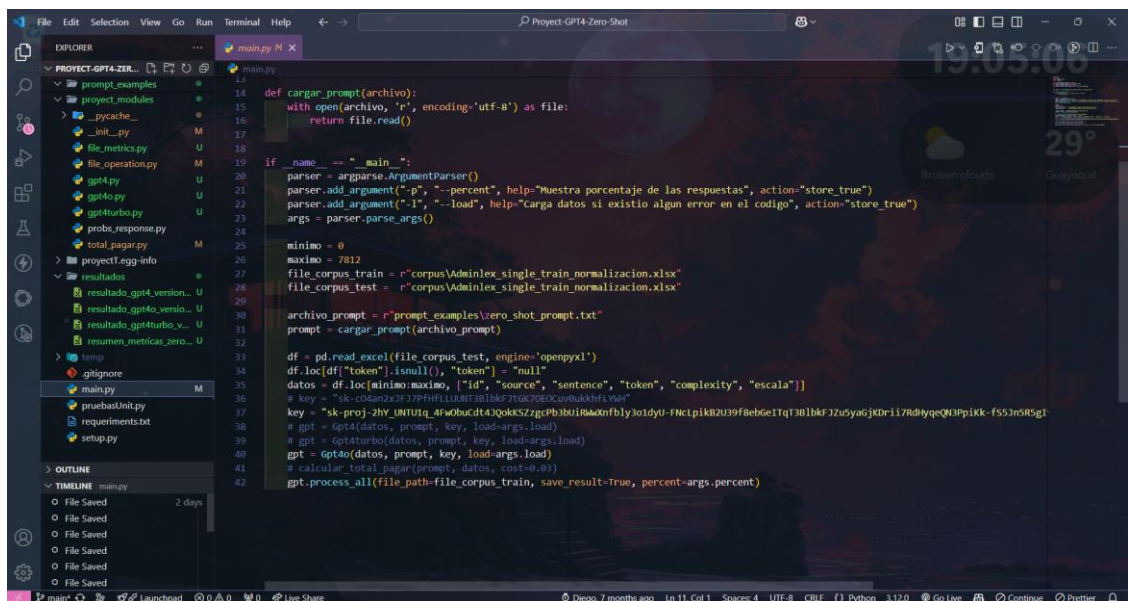
Fig. *Abrir proyecto en vsc desde cmd* **5**



Nota: Elaborado por los autores

Vista del proyecto en vsc

Fig. *Vista del proyecto en vsc* **6**



Nota: Elaborado por los autores

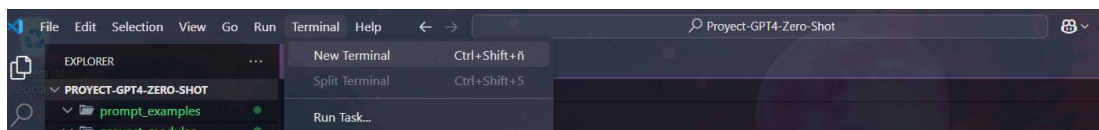
Consideraciones

Antes de ejecutar el proyecto primero debemos tener en cuenta dos cosas, la primera es la clave API key de OpenAI, para procesar los modelos es necesario una clave API con saldo suficiente para procesar un corpus de 7812 registros, para ello se recomienda primero ejecutar el archivo "total_pagar" y lo segundo es descomentar la línea de código del modelo que se desea ejecutar.

Fig.

7

Abrir terminal del proyecto



Nota: Elaborado por los autores

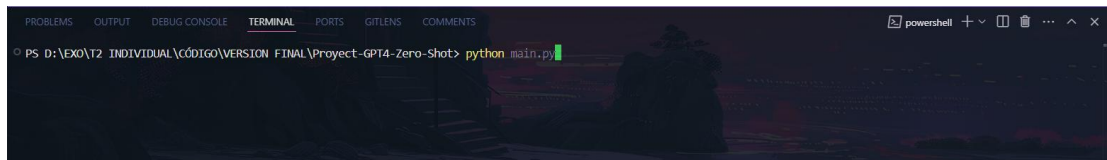
Ejecutar proyecto

Para ejecutar el proyecto y tener los resultados (los excels correspondiente de cada modelo) debemos escribir en la terminal del proyecto 'python main.py'.

Fig.

8

Ejecutar el proyecto



Nota: Elaborado por los autores

Reseña de los autores

Ph.D. JENNY ORTIZ ZAMBRANO



- Analista de Sistemas - Escuela Superior Politécnica del Litoral
- Licenciada en Sistemas e Información - Escuela Superior Politécnica del Litoral
- Diploma Superior en Pedagogía Universitaria - Universidad de Guayaquil
- Diploma Superior en Docencia y Evaluación en la Educación Superior - Universidad de Guayaquil
- Especialista en Gerencia en Educación Superior - Universidad de Guayaquil
- Máster en Docencia y Evaluación de la Educación Superior - Universidad de Guayaquil
- Máster en Desarrollo de Software - Universidad de Granada (España)
- Ph.D. en Tecnologías de la Información y Comunicación - Universidad de Jaén (España)
- Actualmente, docente en la Universidad de Guayaquil.
- Directora de proyectos de vinculación con la sociedad.
- Directora de proyectos de investigación en el campo de la Inteligencia Artificial.
- Autora de varios Libros y Artículos Científicos de alto impacto.
- Expositora en eventos científicos Nacionales e Internacionales.

M.Sc. MIRELLA CARMINA ORTIZ ZAMBRANO



- Abogada de los Tribunales y Juzgados de la República del Ecuador.
 - Licenciada en Ciencias Políticas.
 - Máster en administración pública mención Desarrollo Institucional.
 - Profesora de Segunda Enseñanza de Inglés.
- Docente de la Universidad de Guayaquil, Facultad de Matemáticas, Carrera de Software.
 - Autora de Libros y varios Artículos Científicos de alto impacto
 - Actualmente. Participo como docente investigadora de proyecto competitivo de investigación FCI patrocinado por la Universidad de Guayaquil.

M.Sc. TANIA PERALTA GUARACA



- Ingeniera en Sistemas Computacionales.
- Máster en Docencia y Gerencia en Educación Superior, Universidad de Guayaquil.
- Máster Universitario en Ingeniera de Software y Sistemas Informáticos, Universidad de la Rioja.
- Docente de la Universidad de Guayaquil, Facultad de Ciencias Matemáticas y Físicas,

Carreras de Ingeniería en Sistemas Computacionales y Carrera de Software.

- Gestora de Prácticas Pre-profesionales durante 8 años.
- Actualmente, Gestora General de Vinculación con la Sociedad y Bienestar Estudiantil
- Autora de varios artículos científicos.
- Expositora en Congresos Nacionales.

M.Sc. JESSICA MALENA YEPEZ HOLGUIN



- Máster en Contabilidad y Auditoría, graduada en la Universidad Laica Vicente Rocafuerte de Guayaquil.
- Máster en docencia y Gerencia en Educación Superior - Universidad de Guayaquil.
- Diplomado Superior en docencia y Evaluación en la Educación Superior - Universidad de Guayaquil.
- Ingeniera Comercial - Universidad de Guayaquil.
- Actualmente Docente de contrato ocasional de la Facultad de Ciencias Matemáticas y Física - Universidad de Guayaquil.
- Autora de libros y varios artículos científicos de alto impacto.
- Actualmente, investigadora de proyecto competitivo de investigación FCI patrocinado por la Universidad de Guayaquil.
- Co-evaluadora del área Administrativa - Contable de la Facultad de Ciencias Matemáticas y Física en la Carrera de Ingeniería en Sistemas Computacionales - Universidad de Guayaquil.
- Expositora en Congresos Nacionales e Internacionales.

M.Sc. RICHARD MANOLO CALERO VILLAREAL

- Arquitecto - Universidad de Guayaquil.
 - Máster en Dirección de Proyectos.
 - Profesional con amplia experiencia en diseño arquitectónico, dirección técnica y supervisión de obras residenciales, institucionales y de infraestructura.
- Ha liderado proyectos de construcción aplicando criterios de eficiencia, seguridad y gestión administrativa.
 - Docente universitario con trayectoria en universidades e institutos superiores, aportando formación académica en áreas de arquitectura, construcción y planificación.
 - Actualmente, docente de la Facultad de Ciencias Matemáticas y Físicas de la Universidad de Guayaquil, donde integra su experiencia profesional con la enseñanza, promoviendo innovación, calidad y desarrollo integral en proyectos arquitectónicos.
 - Expositor en espacios académicos y colaborador en iniciativas de mejora institucional y vinculación con la comunidad.



**ING. SCARLET ANNABEL
GUTIERREZ GAVILANEZ**

Ingeniera de software



**ING. IVETTE JUDITH ZAMORA
OLVERA**

Ingeniera de software

ISBN: 978-9942-53-098-1



Compás
capacitación e investigación