

**PONTIFICIA UNIVERSIDAD
CATÓLICA DEL PERÚ**

Escuela de Posgrado



Diseño de un modelo explicativo basado en ontologías
aplicado a un chatbot conversacional

Trabajo de Investigación para obtener el grado académico de Maestro
en Informática con mención en Ciencias de la Computación que
presenta:

Daniel Martin Arteaga Meléndez

Asesor:

Héctor Erasmo Gómez Montoya

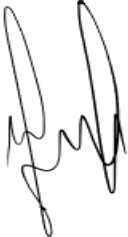
Lima, 2023

Informe de Similitud

Yo, Héctor Erasmo Gómez Montoya, docente de la Escuela de Postgrado de la Pontificia Universidad Católica del Perú, asesor del trabajo de investigación titulado “Diseño de un modelo explicativo basado en ontologías aplicado a un chatbot conversacional”, del autor Daniel Martin Arteaga Meléndez, dejo constancia de lo siguiente:

- El mencionado documento tiene un índice de puntuación de similitud de 17%. Así lo consigna el reporte de similitud emitido por el software *Turnitin* el 24/07/2023.
- He revisado con detalle dicho reporte y el Trabajo de Suficiencia Profesional, y no se advierte indicios de plagio.
- Las citas a otros autores y sus respectivas referencias cumplen con las pautas académicas.

Lugar y fecha: Lima, San Miguel 25 de Julio de 2023

Apellidos y nombres del asesor: Gómez Montoya, Héctor Erasmo	
DNI: 70599170	Firma: 
ORCID: 0000-0002-1338-3392	

Dedicatoria

Este trabajo de investigación lo dedico a mi madre Ana Meléndez por su entrega, enseñanzas y amor que me han permitido llegar a esta etapa y me han formado como profesional.

A mi esposa Gabriela Carrasco por ser mi apoyo incondicional en todo momento y por siempre recordarme mi capacidad y la importancia de mi trabajo.

A mis hermanos Marcel y Claudia, así como a mi padrastro Julver Bahamondes por siempre estar dispuestos a ayudarme con lo que sea necesario para que pueda enfocarme en mi trabajo de investigación.

A mis abuelos Marcelino Sánchez, Sabina Reyes, y Celia Quisel por siempre creer en mí, haberme guiado y dado todo su amor.

A mi padre Dalton Arteaga, por el cariño y preocupación que siempre ha tenido.

Agradecimientos

Agradezco a mi asesor Héctor Gómez por su guía, sus recomendaciones y sus correcciones, sin las que no hubiese podido llegar a esta instancia del trabajo de investigación.

A la Pontificia Universidad Católica del Perú y a mis maestros, por las enseñanzas, vivencias y consejos aprendidos en esta etapa, además de brindarme la posibilidad de realizar un intercambio estudiantil durante un semestre académico.

A la Universidad de Wisconsin – Milwaukee por recibirme y contribuir con mi desarrollo profesional y personal.

A la doctora Susan McRoy por ser mi mentora mientras estuve en la Universidad de Wisconsin – Milwaukee.

Al doctor Samuel Huamán por su orientación en la redacción y sus comentarios en temas de Inteligencia Artificial.

Resumen

Actualmente, la inteligencia artificial es una de las áreas de investigación más importantes para el desarrollo de tecnología en múltiples disciplinas. Aunque ha tenido un crecimiento exponencial en los últimos años, el entendimiento de cómo funciona es mínimo para la mayoría de las personas. En consecuencia de ello, su uso en actividades que implican una toma de decisiones es limitado, lo cual se evidencia en el Reporte 2023 de *Artificial Intelligence Index* [1]. Según este reporte, el cambio porcentual en las respuestas de adopción de la inteligencia artificial por industria y actividad entre el 2021 y 2022 ha sido de -15% y -13% para las actividades de marketing y ventas, y desarrollo de productos y/o servicios, respectivamente. Frente a esto se propone el diseño de un modelo que permita explicar los componentes básicos de un sistema basado en inteligencia artificial a través de un chatbot conversacional en idioma inglés. De este modo, la explicación se brinda en un formato sencillo (texto) y a través de un medio interactivo (conversación). El modelo explicativo se basa en la ontología XAIO, propuesta en este estudio y desarrollada a partir de dos ontologías de aprendizaje de máquina e inteligencia artificial explicable. Haciendo uso de un modelo de generación de lenguaje natural a partir de datos estructurados, el modelo explicativo genera explicaciones en lenguaje natural basadas en el conocimiento descrito en las tripletas de la ontología XAIO. Para evaluar el modelo se implementó un chatbot conversacional que utiliza un modelo de entendimiento de lenguaje natural para identificar intenciones y entidades, a partir de las cuales se realizan las consultas en la ontología que permiten obtener las tripletas. En la evaluación cuantitativa se obtuvo un BLEU promedio de 76.97, lo cual indica un buen desempeño en la tarea de generación de lenguaje natural a partir de datos estructurados. Asimismo, se desarrollaron sistemas de inteligencia artificial explicable con chatbot para la prueba con usuarios y se obtuvo un SUS de 69, indicando una usabilidad por encima del promedio. Finalmente, también se realizó una evaluación cualitativa para obtener las apreciaciones de los participantes acerca de los sistemas, las cuales señalan la

coherencia al momento de responder, la sencillez de las respuestas y la interacción amigable con el chatbot.

Palabras clave: Entendimiento de Lenguaje Natural, Generación de Lenguaje Natural, Modelo Transformer, Ontologías, Sistemas de Inteligencia Artificial Explicable, Chatbot Conversacional.

Abstract

Nowadays, artificial intelligence is one of the most important research areas for the technological development of many disciplines. Although it has grown exponentially in recent years, understanding of how it works is minimal for most people. Consequently, its use in decision making activities is limited, as evidenced in the Artificial Intelligence Index Report 2023 [1]. According to this report, the percentage change in the response of artificial intelligence assimilation by industry and activity between 2021 and 2022 has been -15% and -13% for marketing and sales activities, and product and/or service development, respectively. In view of this, we propose the design of a model to explain the basic components of a system based on artificial intelligence through a conversational chatbot developed in English. Thus, the explanation is provided in a simple format (text) and through an interactive manner (conversation). The explanatory model is based on the XAIO ontology, proposed in this study, and developed from two ontologies of machine learning and explainable artificial intelligence. Using a natural language generation model from structured data, the explanatory model generates natural language explanations based on the knowledge described in the triplets of the XAIO ontology. For evaluation purposes, a conversational chatbot was implemented. This chatbot uses a natural language understanding model to identify intentions and entities. Then it uses ontology queries build from the intentions and entities to get the ontology triplets. Regarding quantitative evaluation, an average BLEU of 76.97 was obtained, which indicates a good performance in the task of natural language generation from structured data. Likewise, explainable artificial intelligence systems were developed with chatbot for user testing and a SUS of 69 was obtained, indicating above-average usability. Finally, a qualitative evaluation was also carried out to obtain the participants' feedback about the systems. They mainly pointed out the coherence at the time of answering, the simplicity of the answers and the friendly interaction with the chatbot.

Keywords: Natural Language Understanding, Natural Language Generation, Transformer Model, Ontologies, Explainable Artificial Intelligence Systems, Conversational Chatbot.

Índice

	Pág.
Resumen	iv
Índice	viii
Índice de Tablas	ix
Índice de Figuras	x
Introducción	1
CAPÍTULO I	
Trabajos Relacionados	5
CAPÍTULO II	
Metodología	8
2.1. Generación del conocimiento	8
2.2. Diseño de la ontología	10
2.3. Implementación del modelo aplicativo	12
2.4. Desarrollo del chatbot conversacional	13
2.5. Implementación de los sistemas de inteligencia artificial explicable	15
CAPÍTULO III	
Evaluación	17
CAPÍTULO IV	
Resultados	20
CAPÍTULO V	
Discusión	22
Conclusiones	24
Recomendaciones	25
Referencias Bibliográficas	26

Índice de Tablas

	Pág.
TABLA I: Preguntas de competencia del sistema de inteligencia artificial explicable	9
TABLA II: Apreciaciones generales de los participantes sobre los sistemas de inteligencia artificial explicable	20

Índice de Figuras

	Pág.
Figura 1: Principales clases y propiedades de la ontología propuesta	11
Figura 2: Flujo para preparación de la respuesta del chatbot desde la pregunta Q_i hasta la explicación en lenguaje natural y_i	14
Figura 3: Sistemas de inteligencia artificial explicable	16



Introducción

El aumento exponencial de los datos, el desarrollo tecnológico de los equipos de cómputo y la propuesta de nuevas técnicas ha permitido la evolución de la inteligencia artificial, especialmente del aprendizaje automático. Esto se evidencia en el número de artículos de inteligencia artificial publicados desde 1980, el cual tiene un comportamiento exponencial en los últimos 12 años [2].

Actualmente, la inteligencia artificial, y específicamente el aprendizaje automático, se emplean en una amplia gama de áreas como medicina [3], educación [4], agricultura [5], negocios [6], auditoría [7], energía [8], etc. Sin embargo, uno de los principales problemas que afrontan estas áreas es la falta de explicabilidad. En el caso del aprendizaje automático, la mayoría de los modelos se ven como cajas negras que la mayoría de las personas no pueden comprender. Esto se convierte en algo crucial cuando uno de estos modelos se va a utilizar para situaciones de toma de decisiones importantes. Tomando un ejemplo del ámbito médico como en [9], si un sistema de apoyo al diagnóstico no es capaz de explicar el proceso de obtención del resultado (o al menos las posibles razones), los médicos nunca confiarán en el sistema.

En este contexto, la inteligencia artificial explicable (XAI por sus siglas en inglés) surgió para hacer frente a esta carencia de los modelos de aprendizaje automático. Esta subárea de la inteligencia artificial pretende ofrecer explicaciones sobre cómo los modelos generan resultados. En ocasiones, las explicaciones se basan en la causa real de los resultados de los modelos y, en otras, en pistas o ideas relacionadas con los resultados (interpretaciones). Esto está estrechamente relacionado con la comprensión del modelo por parte de los usuarios.

A pesar de los avances en el desarrollo de diferentes métodos de explicabilidad, existen algunos retos para la inteligencia artificial explicable. De Brujin et al. [10] describió siete retos principales de la XAI. Uno de los más críticos es el poco o nulo conocimiento para entender las explicaciones acerca de los modelos de inteligencia artificial por parte de los usuarios finales de un sistema. Esto también ha sido indicado por Miller et al. en [11] al indicar que los investigadores en inteligencia artificial desarrollan métodos de explicabilidad que son entendidos

por otros investigadores o personas con cierto nivel de conocimiento de inteligencia artificial, mas no por los usuarios reales.

Una manera de hacer más entendible el contenido de la explicación es utilizar la modalidad adecuada para presentarla. Basados en el hecho que las personas generalmente dan explicaciones verbales, muchas investigaciones se han enfocado en que los sistemas de inteligencia artificial explicable brinden explicaciones textuales. Sin embargo, gran parte de las explicaciones textuales se elaboran a partir de un conjunto de frases predefinidas o mediante el completado de plantillas preestablecidas. Ante esta situación, algunos trabajos vienen proponiendo el uso de Procesamiento de Lenguaje Natural para generar explicaciones textuales más eficientes y comprensibles para cualquier usuario [12], [13].

Aunque una explicación textual resulta cómoda para la mayoría de las personas, algunas veces son difíciles de entender, especialmente cuando son textos muy largos o dónde se requiere contexto y/o retroalimentación. Para este tipo de explicaciones, una conversación es un medio interactivo en el que se expresan ideas en forma de textos relativamente cortos según un contexto entendible por los participantes y dónde existe una retroalimentación continua. En base a ello, se podría emplear un chatbot conversacional como la interfaz explicativa de un sistema basado en inteligencia artificial.

Un aspecto importante cuando se utiliza un chatbot es gestionar la información o conocimiento a partir del cual se generarán las respuestas. Muchos chatbots de dominio cerrado utilizan bases de datos, puesto que para la mayoría de las aplicaciones se trabaja con información tabular; sin embargo, para el caso de generar explicaciones en lenguaje natural que sean entendibles se requiere una mejor forma de gestionar la información de las explicaciones. Para esta finalidad existen las bases de conocimiento y la Web Semántica provee de una base de conocimiento ideal para gestionar los conceptos y las relaciones que se usan para describir cierto dominio denominada ontología. Aunque algunos trabajos [14], [15] han usado ontologías para gestionar información base de modo que un chatbot pueda elaborar respuestas, existen pocas ontologías que permitan abarcar los conceptos necesarios para la generación de explicaciones sobre datos, modelos y sistemas de inteligencia artificial.

En base a las limitaciones de los sistemas de inteligencia artificial explicable y los pocos mecanismos para gestionar información de las explicaciones, se identifica una carencia de sistemas de inteligencia artificial explicable eficientes para una variedad de situaciones conformada por diferentes datos, modelos y sistemas de inteligencia artificial. Como se ha comentado en el Resumen y tomando de evidencia datos del Reporte 2023 de *Artificial Intelligence Index* [1], este problema genera desconfianza en los sistemas basados en inteligencia artificial debido a la dificultad de interpretación de las explicaciones, lo que a su vez limita el desarrollo y utilización de la inteligencia artificial en aplicaciones para toma de decisiones.

Este artículo propone emplear un modelo explicativo que genere explicaciones en lenguaje natural automáticamente a partir de información de explicabilidad gestionada en una ontología y cuyas explicaciones se brinden a través de un chatbot conversacional. La generación de explicaciones en lenguaje natural se realiza a través de un modelo *transformer* propuesto en [16] y que ha sido previamente preentrenado con múltiples conjuntos de datos etiquetados para diferentes tareas de generación de lenguaje natural. El modelo fue ajustado para la tarea de generación de textos a partir de datos estructurados con un conjunto de datos elaborado a partir de tripletas extraídas de la ontología propuesta, las cuales contienen el conocimiento necesario para elaborar las explicaciones en lenguaje natural. Este conocimiento se creó a partir del desarrollo de algunos modelos y permite responder un conjunto de preguntas prototipo sobre un sistema basado en inteligencia artificial.

La evaluación del modelo explicativo se llevó a cabo a través de la métrica para la tarea de generación de lenguaje natural denominada BLEU [17]. Sin embargo, en algunas situaciones esta métrica no es suficiente para sustituir la interpretación y el juicio humano. Por esta razón se evaluó todo el sistema de inteligencia artificial explicable, es decir, el modelo explicativo como parte del chatbot conversacional, mediante una evaluación humana que consistió en pruebas con usuarios de poco o nulo conocimiento acerca de inteligencia artificial. Las pruebas plantearon escenarios en los que los participantes debían emplear sistemas basados en inteligencia artificial. Después de la interacción con los sistemas se les solicitó completar un cuestionario de entrevista para

recopilar sus apreciaciones del sistema y una encuesta con preguntas que permitió medir qué tan comprensibles han sido los sistemas.

El aporte principal del trabajo radica en la generación de explicaciones en lenguaje natural a partir de una ontología para explicabilidad de modelos de inteligencia artificial. Además de ello se propone como interfaz explicativa la interacción con un chatbot conversacional.

Este artículo se organiza de la siguiente forma: los trabajos relacionados se presentan en la sección II; la metodología seguida se describe en la sección III, detallando el diseño de la ontología para explicabilidad de modelos de inteligencia artificial, la implementación del modelo explicativo y el desarrollo del chatbot conversacional; la sección IV indica el proceso de evaluación del modelo explicativo y del sistema de inteligencia artificial explicable completo; los resultados se comentan y se discuten en la sección V y sección VI respectivamente; finalmente, la sección VII presenta las conclusiones de esta investigación.

CAPÍTULO I

Trabajos Relacionados

Por mucho tiempo se pensó que los métodos de explicabilidad debían de obtener como resultado únicamente valores numéricos o gráficas interpretables por los especialistas en inteligencia artificial. Actualmente, los métodos de explicabilidad permiten obtener explicaciones más intuitivas para la mayoría de los usuarios a través de texto. Tal y como lo describe Papastratis en [18], la categorización de métodos de explicabilidad de acuerdo con su modalidad son: métodos de interpretación visual, explicaciones textuales y explicaciones matemáticas o numéricas.

Las explicaciones textuales han tenido un gran interés debido a los recientes avances en Procesamiento de Lenguaje Natural. Lei et al. [12] desarrollaron un método para extraer partes de los textos de entrada como justificaciones o argumentos para posteriormente señalar los resultados de las predicciones en la tarea de análisis de sentimientos. Por otro lado, Donadello et al. [13] resalta la importancia del componente de generación de lenguaje natural de su solución para brindar explicaciones contextuales efectivas y persuasivas en el dominio de recomendación de una dieta saludable.

La interfaz por la que se brinda las explicaciones es tan importante como las mismas explicaciones. Esto se evidencia en el trabajo de Galitsky [19] donde propuso una interfaz conversacional para registrar las decisiones de un sistema de aprendizaje automático que enumera los pasos seguidos y las características empleadas para llegar a una decisión determinada. Debido a la interacción con el agente conversacional, el conocimiento del sistema para registrar las decisiones mejora y permite proporcionar respuestas más completas. Además, las personas que utilizan la interacción conversacional con el sistema desarrollan una mayor confianza que las personas que interactúan con la explicación textual tradicional basada en párrafos o informes. De igual forma, Sun et al. [20] exploraron los efectos del diálogo interactivo para dar explicaciones sobre los pronósticos de sistemas de diagnóstico en línea. Los investigadores compararon las explicaciones en formato de diálogo con las explicaciones textuales estáticas y con sistemas sin explicaciones.

Algunos autores han estudiado el uso de representaciones del conocimiento como apoyo a la generación de explicaciones. Celino [21] describe la importancia de la representación del conocimiento y el razonamiento en la XAI. Señala que la XAI puede aprovechar toda la experiencia de anotadores humanos y las bases de conocimiento y grafos de conocimiento de múltiples dominios. Un ejemplo de esto fue el trabajo de Basu et al. [22], en donde los autores presentan un enfoque para modelar y gestionar la información necesaria para la explicación a través de una base de datos relacional. Se muestra cómo se pueden producir distintos tipos de explicaciones mediante las consultas adecuadas en esta base de datos. Asimismo, Inoue [23] afirmó que las tecnologías de la web semántica proporcionan herramientas para razonar en recursos de conocimiento como las ontologías y, por tanto, se han aplicado para hacer explicable el aprendizaje automático. Además, desarrolló una ontología que representa experimentos de aprendizaje automático explicables.

El uso de sistemas agentes conversacionales para brindar explicaciones ha sido estudiado previamente por algunos investigadores. D. Gunning et al. [24] menciona que un equipo de Raytheon BBN Technologies (conformado por investigadores del Instituto Tecnológico de Georgia, del Instituto Tecnológico de Massachusetts y de la Universidad de Texas) se encontraban desarrollando un sistema para responder cualquier pregunta sobre datos multimedia y proporcionar explicaciones interactivas y explorables de cómo se elaboran las respuestas. Por otro lado, Khurana et al. [25] propusieron ChatrEx, un chatbot con una interfaz explicable capaz de explicar el funcionamiento subyacente durante un problema de entendimiento del chatbot. ChatrEx-VINC ofrece explicaciones paso a paso basadas en ejemplos visuales en el contexto de la ventana de chat, mientras que ChatrEx-VST ofrece explicaciones como un recorrido visual superpuesto en la interfaz de la aplicación.

Nguyen et al. [26] proponen el uso de un agente conversacional para explicabilidad de modelos de aprendizaje automático, con un enfoque similar al que proponemos en este trabajo; sin embargo, la principal diferencia radica en la fuente de información y la forma en cómo se generan las explicaciones. Nguyen et al. emplean la información obtenida como resultado de los métodos de explicabilidad como SHAP o DICE; la generación de explicaciones textuales

se basa en plantillas con vocabulario propio del dominio de los datos. Por otro lado, en este trabajo proponemos el uso de una ontología diseñada para gestionar información de los diferentes datos, modelos, sistemas, entre otros aspectos; asimismo, el método para generar explicaciones es a través de un modelo *transformer* ajustado para la tarea de convertir textos en formato de datos a textos en lenguaje natural, usando como datos de entrada las tripletas extraídas de la ontología.

CAPÍTULO II

Metodología

Esta sección describe los principales componentes de la metodología seguidos en esta investigación, los cuales corresponden a la generación del conocimiento para brindar explicaciones, el diseño de la ontología, la implementación del modelo explicativo, el desarrollo del chatbot conversacional y la implementación de los sistemas de inteligencia artificial explicable.

2.1. Generación del conocimiento

Con la finalidad de gestionar la información necesaria para crear explicaciones en lenguaje natural de modelos de inteligencia artificial se ha propuesto el uso de una ontología. Una etapa previa al diseño de la ontología ha sido la identificación de las clases, propiedades y los individuos que forman parte de esta.

Para llevar a cabo esto se empezó implementando diferentes tipos de modelos de inteligencia artificial en las tareas de regresión y clasificación a partir de una variedad de conjuntos de datos. Los modelos implementados son: regresión lineal, regresión logística, árboles de decisión, K vecinos más cercanos (KNN), bosques aleatorios (RF) y máquinas vectoriales de soporte (SVM). Los conjuntos de datos empleados fueron versiones preprocesadas de: *Bike Sharing Dataset* [27], *Campus Recruitment Dataset* [28], *Pima Indians Diabetes Database* [29], y *Heart Disease Dataset* [30]. La implementación de estos modelos comprende los pasos de preparación de los datos o preprocesamiento, la creación de los modelos, el entrenamiento y la evaluación de estos. Posteriormente, se aplicó el método de importancia de variables tras permutación (*Permutation feature importance*) [31] para poder respaldar los resultados del sistema.

El siguiente paso fue la elaboración de un conjunto de preguntas que requieren explicación acerca del sistema basado en inteligencia artificial, los datos, los modelos, el proceso de implementación y los resultados. Estas preguntas se plantearon pensando en los usuarios finales de los sistemas basados en

inteligencia artificial, los cuales poseen limitado conocimiento de este campo. Además, debido a la disposición de más información en inglés acerca de los conocimientos en las ontologías empleadas y los textos para entrenar los modelos de entendimiento de lenguaje natural y generación de lenguaje natural, se ha optado por emplear este lenguaje. En la Tabla I se muestran algunos ejemplos de las preguntas que el sistema de inteligencia artificial explicable puede responder.

TABLA I: Preguntas de competencia del sistema de inteligencia artificial explicable

Intenciones	Preguntas
Objetivo del sistema	What is the system's goal?
Funcionamiento del sistema	How does this system work?
Tarea de inteligencia artificial que realiza	What is classification?
Tipo de modelo de inteligencia artificial	What kind of model does the system use?
Evaluación del modelo	How was the model evaluated?
Descripción de la métrica de evaluación	What is the F1 score?
Interpretación de la métrica de evaluación	How do you interpret this measure?
Conjunto de datos base del sistema	What data is the system based on?
Características del conjunto de datos	What is the number of training instances of the dataset?
Variables de entrada, salida e hiperparámetros	What are the hyperparameters of the model?
Descripción de las variables	What is the meaning of resting blood pressure?
Tipos de variables	What type of variable is temperature?
Características de las variables	What is the standard deviation of age?
Importancia de las variables	What is the importance of glucose concentration?
Variable más importante	What is the most important feature?
Variable menos importante	What is the least significant variable?

2.2. Diseño de la ontología

Con las clases, propiedades e individuos identificados se buscaron ontologías en los dominios de aprendizaje automático y explicabilidad. La ontología *ML Schema* (MLS) [32] provee de un conjunto de clases y propiedades para representar e intercambiar información sobre algoritmos de aprendizaje automático y de minería de datos, conjuntos de datos y experimentos. Por otro lado, la ontología *Explanation Ontology* (EO) [33] fue diseñada para captar los aspectos de las explicaciones que están relacionados con su generación desde la perspectiva de un sistema y para un usuario final. La ontología diseñada a partir de estas dos se ha denominado *eXplainable Artificial Intelligence Ontology* (XAIO). En futuros estudios continuaremos realizando modificaciones a la ontología de acuerdo con las necesidades de explicabilidad de los usuarios, por esta razón, la ontología aún no será publicada.

El primer paso de esta etapa fue integrar ambas ontologías a partir de las clases y propiedades en común y aquellas que son similares. En este paso se empleó el software Protégé en su versión 5.5.0. Debido a que la ontología EO posee una estructura jerárquica más extensa de clases y propiedades, fue usada como base jerárquica de la ontología propuesta. Para el caso de las clases y propiedades en común se dio prioridad a las de la ontología EO; sin embargo, si las mismas existían en la ontología MLS se agregaron los axiomas faltantes.

En el caso de las clases y propiedades que eran similares se crearon, modificaron y eliminaron los axiomas para establecer únicamente las relaciones de importancia para la generación de las explicaciones. Las clases y propiedades de la ontología MLS que no estaban presentes en la ontología EO se agregaron siguiendo la estructura jerárquica de la última. También fue necesario crear algunas clases y propiedades adicionales para poder elaborar adecuadamente ciertas explicaciones. La Fig. 1 muestra las clases y propiedades más importantes de la ontología propuesta.

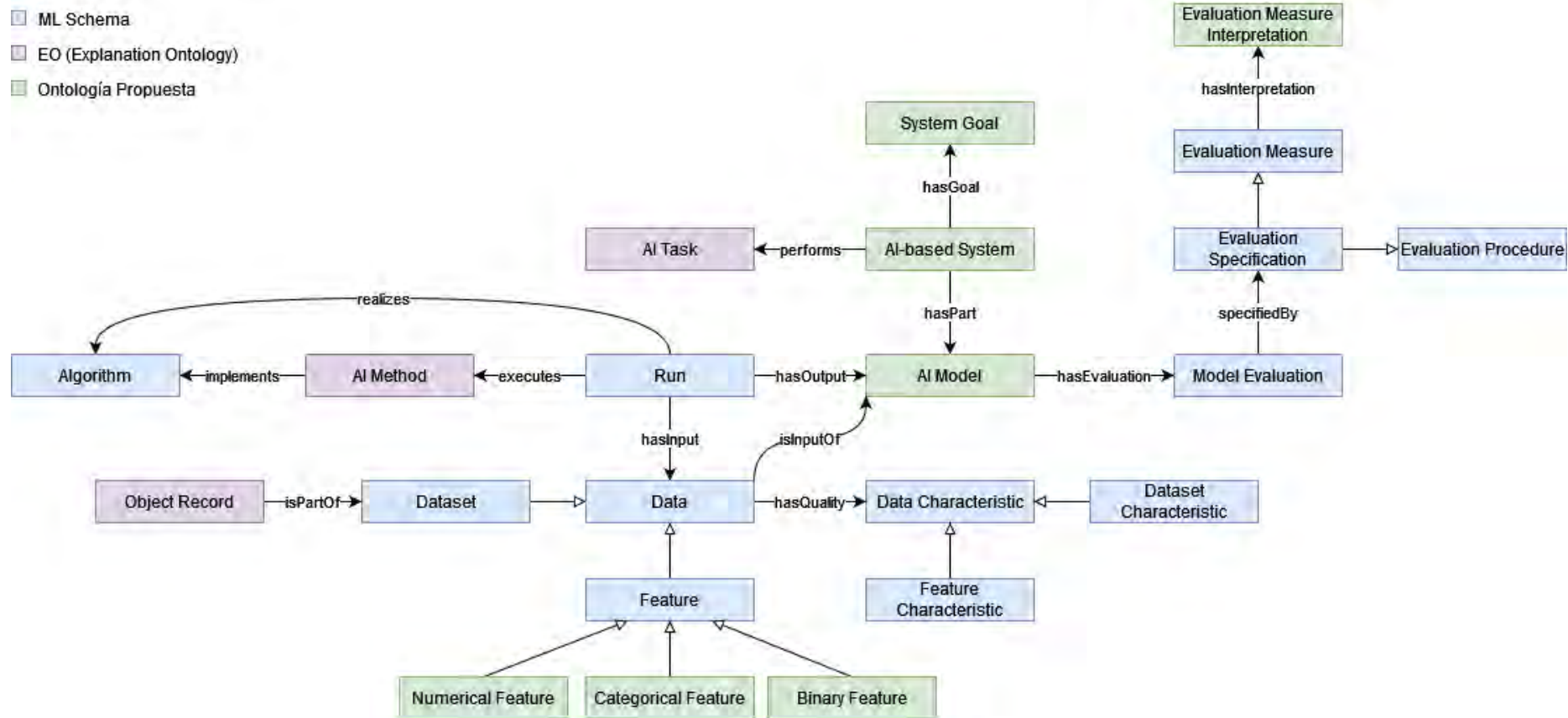


Fig. 1: Principales clases y propiedades de la ontología propuesta.

Una vez que se tenían las clases y propiedades definidas, se procedió a crear los individuos y a asignar valores a algunos de ellos a través de las propiedades de datos. Estos individuos corresponden a los diferentes datos, modelos, procedimientos de evaluación, sistemas basados en inteligencia artificial e información adicional relacionada que se generó durante la implementación de los modelos de inteligencia artificial descrita en la subsección anterior. La creación de algunos de los individuos se realizó de forma manual a través del software Protégé. El resto de los individuos se agregó de forma automática empleando el lenguaje de programación Python y la librería “owlready2”.

Finalmente, se procedió a generar todas las tripletas obtenidas a partir de las clases, individuos y propiedades de la ontología. Estas tripletas servirán para poder conformar el corpus con el que se ajustará el modelo de generación de texto a partir de datos estructurados.

2.3. Implementación del modelo aplicativo

Para implementar el modelo explicativo se ha llevado a cabo la conformación del corpus para la tarea de generación de texto a partir de datos estructurados y el ajuste del modelo.

El corpus se compone de instancias cuyas entradas x_i son los textos en formato tripleta que poseen la información de los datos estructurados, y que tienen como salidas y_i los textos en formato oración. Las entradas de las instancias pueden ser simples o compuestas. Las entradas simples corresponden a cada tripleta de la ontología XAIO y sus textos tiene la estructura *sujeto|propiedad|objeto*. Las entradas compuestas se generaron a partir de múltiples tripletas que expresan una idea compuesta y que poseen una clase o individuo en común. Una posible estructura de un texto de entrada compuesta es por ejemplo *sujeto₁|propiedad₁|sujeto₂ [SEP] sujeto₂|propiedad₂|objeto₂*. Por otro lado, las salidas de las instancias son textos formados por una oración y por más oraciones para las salidas simples y compuestas respectivamente. Todas las entradas de las instancias se generaron automáticamente con un script en Python, mientras que las salidas se tuvieron que redactar una a una. En total se obtuvieron 1845 ejemplos de entrenamiento y 206 ejemplos de prueba.

El siguiente paso es ajustar el modelo de generación de explicaciones en lenguaje natural. Se ha utilizado el modelo denominado MVP (*Multitask superVised Pre-training*), que posee una arquitectura *transformer* codificador – decodificador. Este modelo ha sido previamente preentrenado en dos etapas: entrenamiento del modelo usando diferentes conjuntos de datos para 7 diferentes tareas de generación de lenguaje natural, entre las que está la tarea de generación de texto a partir de datos estructurados; y entrenamiento del modelo para una tarea específica de generación de lenguaje natural empleando una “instrucción” [16]. En esta investigación se ha empleado el modelo MVP de Hugging Face que en la segunda etapa fue preentrenado para la tarea de generación de textos a partir de datos estructurados. Este modelo ha sido ajustado con el conjunto de datos que se compuso a partir de las tripletas de la ontología XAIO y las explicaciones redactadas manualmente. La configuración para este ajuste corresponde a $2e-5$ de ratio de aprendizaje, 8 ejemplos por lote (*batch size*) tanto para entrenamiento como evaluación, 0.01 de decaimiento de pesos (*weight decay*), 5 épocas de entrenamiento y evaluación al final de cada época.

2.4. Desarrollo del chatbot conversacional

El chatbot conversacional se ha desarrollado para poder brindar las explicaciones en formato de texto, siguiendo el flujo descrito en la Fig. 2. El primer paso en la elaboración de la respuesta es el entendimiento de la pregunta Q_i . Para ello se ha implementado un módulo de entendimiento de lenguaje natural capaz de obtener la intención I_i y las entidades E_i de la pregunta. El siguiente paso es consultar por la información requerida en la ontología XAIO. Las consultas q_i se encuentran definidas de acuerdo con la intención detectada I_i y las entidades reconocidas E_i en el paso previo. Las consultas q_i se realizan en lenguaje SPARQL y fueron diseñadas para devolver tripletas x_i de la forma en que se mencionó en la subsección anterior. El último paso es emplear el modelo explicativo previamente descrito (como parte del módulo de generación de lenguaje natural) para convertir el resultado de la consulta en formato tripleta x_i (datos estructurados) a una explicación en formato de lenguaje natural y_i . En

la parte intermediaria de este chatbot se encuentra el administrador de la conversación.

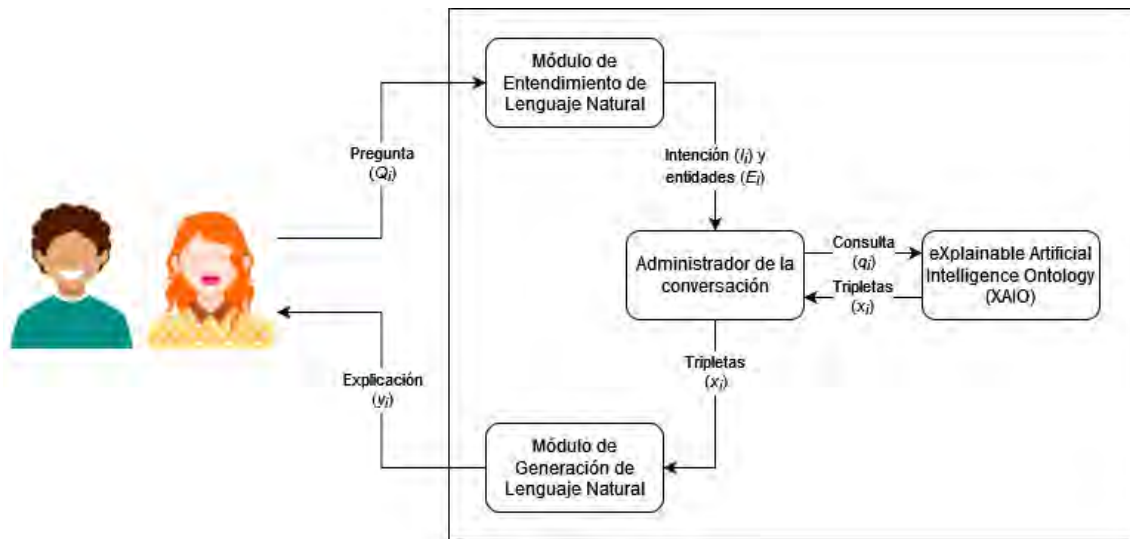


Fig. 2: Flujo para preparación de la respuesta del chatbot desde la pregunta Q_i hasta la explicación en lenguaje natural y_i .

El módulo de entendimiento de lenguaje natural se basa en el modelo JoinBert [34], extendiendo el modelo preentrenado Bert para realizar de manera conjunta la clasificación de las intenciones y el reconocimiento de las entidades nombradas. En esta arquitectura, se utiliza la aparición del token (“[CLS]”) para representar la secuencia completa de entrada, la cual es pasada como entrada a una red densa para la clasificación de intenciones. Por otro lado, para el reconocimiento de la entidad nombrada, la salida de cada token de entrada individual es pasado a la misma capa totalmente conectada para generar la entidad del token. Dado que cada token BERT toma un enfoque tokenización de subpalabras o *WordPiece Embedding*, del mismo modo que en [35] se asigna la etiqueta de la entidad solo a la primera subpalabra ignorando las siguientes subpalabras.

El conjunto de datos para el entrenamiento de este modelo se crea a partir de una lista base de 16 preguntas que el sistema puede responder y que algunos ejemplos se indicaron en la Tabla 1. Estas preguntas fueron aumentadas mediante el parafraseo de las mismas. El conjunto de datos es dividido en un 80% para el entrenamiento y un 20% para el conjunto de pruebas. Además, se

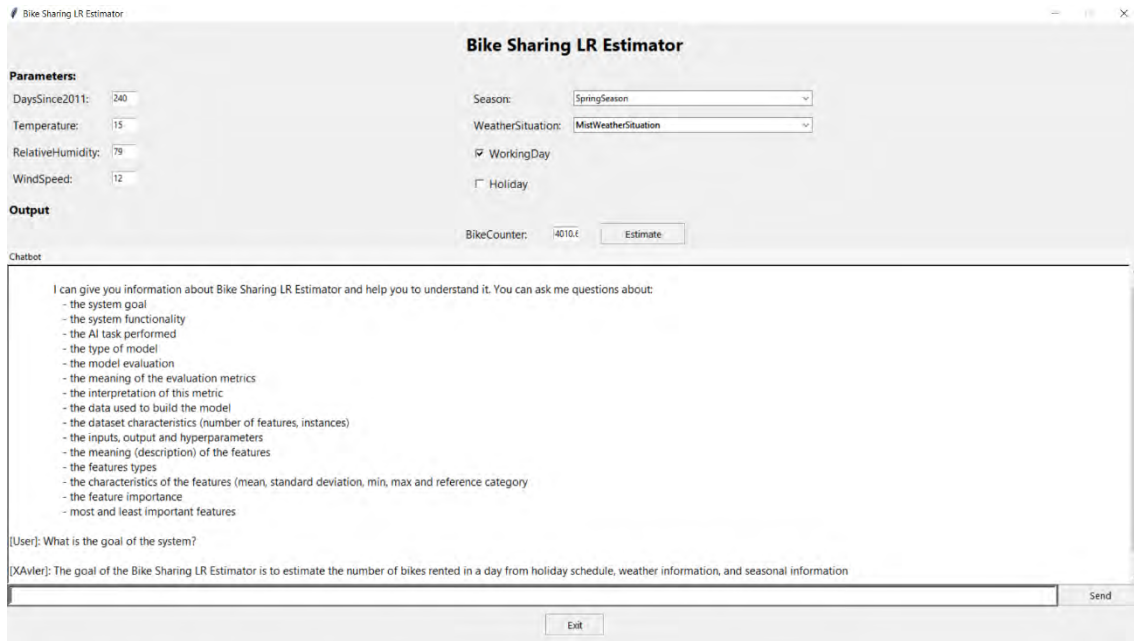
realiza el aumento de los ejemplos, realizando en cada partición el reemplazo de palabras clave por sinónimos y el reemplazo de las entidades por nuevas entidades obtenidas tras la incorporación de todos los modelos y conjuntos de datos empleados. Así, finalmente se obtuvo un total de 1200 ejemplos en el conjunto de entrenamiento y 300 ejemplos en el conjunto de prueba.

2.5. Implementación de los sistemas de inteligencia artificial explicable

De acuerdo con el marco explicativo propuesto por la Agencia de Proyectos de Investigación Avanzados de Defensa de los Estados Unidos [24], un sistema de inteligencia artificial explicable es capaz de desarrollar la tarea, obtener un resultado y proporcionar la explicación que justifique el resultado; para ello, el sistema debe conformarse por el modelo de inteligencia artificial, el modelo explicativo, y la interfaz de explicación. En este trabajo se han implementado dos sistemas de inteligencia artificial explicable para desarrollar las pruebas.

El primer sistema de inteligencia artificial explicable se enfoca en la tarea de regresión usando un modelo de regresión lineal múltiple y el conjunto de datos *Bike Sharing Dataset* [27]. El segundo sistema realiza la tarea de clasificación mediante un modelo de máquina vectorial de soporte y el conjunto de datos *Heart Disease Dataset* [30]. El modelo explicativo de ambos sistemas es el propuesto en este trabajo y la interfaz de explicación es un chatbot conversacional.

Se desarrolló interfaces de usuario sencillas usando la librería “Tkinter” de Python para que el usuario pudiera configurar las variables de entrada de los sistemas y poder obtener los resultados. Además, en la parte inferior de los sistemas se muestra el chatbot conversacional. La Fig. 3 muestra las interfaces de estos dos sistemas de inteligencia artificial explicable.



(a) *Bike Sharing LR Estimator*: sistema para la tarea de regresión



(b) *Heart Disease SVM Predictor*: sistema para la tarea de clasificación

Fig. 3: Sistemas de inteligencia artificial explicable.

CAPÍTULO III

Evaluación

La evaluación del modelo explicativo se ha realizado en dos contextos. Primero, se ha evaluado durante el desarrollo de este a través de la métrica BLEU [17] usada para las tareas de generación de texto. Sin embargo, dado que el objetivo del modelo explicativo es generar textos que permitan la interpretación de los resultados del sistema basado en inteligencia artificial y el entendimiento del sistema y los datos, una evaluación con juicio humano es necesaria. En ese sentido, el segundo escenario es la evaluación de todo el sistema de inteligencia artificial explicable (modelo de inteligencia artificial, modelo explicativo y chatbot) mediante una evaluación humana a través de pruebas con usuarios de poca o nula experiencia en inteligencia artificial.

Para esta evaluación se ha empleado un enfoque metodológico mixto. Por un lado, se ha usado un enfoque cualitativo para conocer las percepciones de los usuarios finales acerca de los sistemas de inteligencia artificial explicable; y por el otro lado, el enfoque cuantitativo se ha empleado de dos formas: primero, para medir el rendimiento del modelo generativo en la tarea de elaborar explicaciones en lenguaje natural a través de la métrica BLEU; y en segundo lugar, para medir la relación entre el entendimiento de los sistemas basados en inteligencia artificial percibidos por los usuarios y la utilización del modelo explicativo propuesto como parte de los sistemas de inteligencia artificial explicable.

La evaluación cuantitativa del modelo explicativo a través de la métrica BLEU se realizó durante el desarrollo del modelo haciendo uso del conjunto de datos de prueba. Esta métrica fue desarrollada para evaluar los textos generados por un sistema de traducción automática frente a textos de referencia. Cuando hay una correspondencia perfecta entre los textos se obtiene un valor BLEU de 1 (o 100 en porcentaje); mientras que cuando no existe correspondencia alguna entre los textos, el valor BLEU es 0. Las principales razones para emplear esta métrica son su rapidez y costo de cómputo, su sencillez de entendimiento, la cualidad de ser independiente del lenguaje y que en muchos casos su valor está correlacionado con la evaluación humana.

Para la evaluación a través de las pruebas de usuario, hasta la redacción de este trabajo, se ha contado con la participación de 20 usuarios. La media de la edad de los participantes es 34.6 años con una desviación estándar de 11.54 años. Para esta prueba todos los participantes poseen poca experiencia en inteligencia artificial. El objetivo de la prueba con usuarios es validar la idea de que el modelo explicativo, a través del chatbot conversacional, puede ayudar a los usuarios interpretar y/o comprender los sistemas basados en inteligencia artificial.

Los sistemas de inteligencia artificial explicable mencionados en la sección anterior fueron diseñados para poder hacer dos tareas al mismo tiempo: usar el sistema basado en inteligencia artificial para obtener las salidas (demanda de bicicletas a rentar por día e indicador de riesgo alto de padecer una enfermedad del corazón), y realizar las preguntas que desee al chatbot conversacional. Tras finalizar la interacción con los sistemas se procede con la segunda evaluación cuantitativa, la cual consiste en responder una encuesta con 10 ideas u opiniones que permitan valorar qué tan comprensibles han sido estos sistemas debido a las explicaciones dadas por el chatbot conversacional. Las ideas u opiniones se han adaptado de la encuesta de la Escala de Usabilidad del Sistema (SUS) para la explicabilidad a través de un chatbot conversacional. Las posibles respuestas a la encuesta corresponden a los niveles de desacuerdo o acuerdo con las ideas que van desde 1 (totalmente en desacuerdo) a 5 (totalmente de acuerdo). Para calcular la métrica de evaluación se seguirán directrices de la encuesta SUS. Normalmente, un valor superior a 68 indica que la usabilidad del sistema está por encima del promedio. Las ideas u opiniones de esta encuesta fueron las siguientes:

- 1) Me gustaría utilizar chatbots que explican sistemas basados en inteligencia artificial con más frecuencia.
- 2) La interacción con el chatbot me resultó compleja a efectos de explicabilidad.
- 3) Creo que la comunicación con el chatbot fue fácil y clara.
- 4) Creo que necesitaría soporte técnico para entender las explicaciones dadas por el chatbot.
- 5) Las respuestas del chatbot me parecieron bien elaboradas.
- 6) Creo que hay incoherencias en las respuestas proporcionadas.

- 7) Creo que la mayoría de la gente aprendería sobre el sistema a través de las explicaciones del chatbot.
- 8) Las explicaciones del chatbot me han parecido muy engorrosas.
- 9) Me sentí muy seguro con las explicaciones del chatbot.
- 10) Necesito aprender más cosas antes de poder entender las respuestas del chatbot.

La evaluación cualitativa se ha realizado a través de un cuestionario de entrevista para recopilar las apreciaciones de los usuarios finales luego del uso de los sistemas de inteligencia artificial explicable. Este cuestionario se compone de las siguientes preguntas:

- ¿Cómo le ha parecido el desempeño del chatbot para explicar el sistema basado en Inteligencia Artificial?
- ¿Cuáles son las características por mejorar del sistema?
- ¿Cuáles ha sido las explicaciones más útiles del sistema?
- ¿Qué otras explicaciones le gustaría que el sistema pueda brindar?

CAPÍTULO IV

Resultados

La métrica BLEU obtenida para el conjunto de prueba tuvo un valor promedio de 76.97, un valor mínimo de 61.41, y un valor máximo de 98.06. El valor promedio del BLEU representa un buen desempeño del modelo explicativo en generar textos comparables a los textos de referencia. Esto puede deberse a que las formas de brindar las explicaciones pueden ser no muy variables para las preguntas de competencia planteadas.

Por otro lado, en la evaluación cuantitativa a través de la métrica SUS, se obtuvo un valor promedio de 70.35 con una desviación estándar de 8.71. Esto se puede interpretar como que la usabilidad del sistema de inteligencia artificial explicable mediante el chatbot conversacional está por encima del promedio.

En cuanto a la evaluación cualitativa, las apreciaciones generales de los participantes respecto a los sistemas de inteligencia artificial explicable en las preguntas del cuestionario para entrevista se resumen en la Tabla II.

TABLA II: Apreciaciones generales de los participantes sobre los sistemas de inteligencia artificial explicable

Preguntas	Apreciaciones generales
¿Cómo le ha parecido el desempeño del chatbot para explicar el sistema basado en Inteligencia Artificial?	Eficiente, claro, conciso e interesante. Resaltaron la coherencia al momento de responder.
¿Cuáles son las principales bondades del sistema?	La sencillez en generar las respuestas, la interacción amigable a través de chatbot y que permitió afianzar algunos conceptos de inteligencia artificial.
¿Cuáles son las características por mejorar del sistema?	La variabilidad en que se puede realizar las preguntas, el vocabulario empleado para explicar los conceptos, y la base del conocimiento.
¿Cuáles ha sido las explicaciones más útiles del sistema?	La descripción de las variables, el significado de las métricas de evaluación, las características de las variables, el objetivo del sistema, y el conjunto de datos empleados.

¿Qué otras explicaciones le gustaría que el sistema pueda brindar?	Impacto de los modelos y conceptos relacionados con el negocio. También indicaron profundizar en la descripción de algunos términos técnicos.
--	---

CAPÍTULO V

Discusión

El valor de la métrica BLEU para el conjunto de prueba indica una performance superior en la generación de explicaciones que lo reflejado por el valor de la métrica SUS, obtenida a partir de la encuesta de usabilidad del sistema de inteligencia artificial explicable mediante el chatbot conversacional. El motivo de esta observación se obtiene de las apreciaciones generales de los participantes respecto a la pregunta N°3 (¿Cuáles son las características por mejorar del sistema?) que se muestran en la Tabla II, donde los participantes indican que la variabilidad para realizar las preguntas es limitada. Si bien es cierto el desempeño del módulo de entendimiento de lenguaje natural es independiente de la del módulo de generación de lenguaje natural, el hecho de limitar las posibles formas en que se realiza una pregunta influye en las posibles formas de responderla. Por ejemplo, para el entrenamiento del módulo de lenguaje natural una de las intenciones estudiadas fue “describir la evaluación del modelo” y en ese sentido se elaboraron preguntas como: “¿Cómo fue evaluado el modelo?”, “¿Cómo se ha llevado a cabo la evaluación del modelo?”, y “¿Cuál fue el desempeño del modelo en la evaluación?”. Frente a estas preguntas, se planteó que el sistema responda indicando el proceso de evaluación y la métrica de evaluación usada, así como el valor obtenido para dicha métrica. Sin embargo, para esta intención no se plantearon preguntas ni respuestas exclusivamente acerca de la métrica de evaluación como “¿Cuál ha sido la métrica utilizada?”. Lo que generó errores en la conversación como las que se muestran en el siguiente extracto de la conversación con uno de los usuarios, en donde el chatbot no pudo entender al usuario en las dos primeras preguntas y no fue hasta que cambió la forma en que preguntó que el chatbot pudo responder adecuadamente.

User: What are the evaluation metrics?

XAvler: Sorry, I cannot understand the question. Probably the error is: "no evaluation measure have been mentioned"

User: What metrics do you use?

XAvler: Sorry, I cannot understand the question. Probably the error is: "no evaluation measure have been mentioned"

User: How was this model evaluated?

XAvler: The evaluation of the multiple linear regression model has an evaluation procedure and evaluation measure. The evaluation procedure was five-fold cross-validation and the evaluation measure was adjusted R-squared. The value of the adjustment measure was 0.767

Además de ello, es importante tener en cuenta que algunos errores en el módulo de entendimiento de lenguaje natural se deben a que los participantes no son hablantes del lenguaje inglés y sus preguntas no fueron correctamente redactadas. Estos errores de entendimiento se tradujeron en la identificación errónea de intenciones y entidades, llevando a una generación de explicaciones errónea. Un ejemplo de esta interacción se muestra en el siguiente extracto de una conversación.

User: How is the functionality?

XAvler: The evaluation of the decision tree regressor model has an evaluation procedure and evaluation measure. The evaluation procedure was five-fold cross-validation and the evaluation measure was adjusted R-squared. The value of the adjustment measure was 0.753

User: What are characteristics dataset?

XAvler: -

User: What are the feature of the system?

XAvler: This system is trained to estimate the number of bikes rented in a day from holiday schedule, weather information, and seasonal information

Conclusiones

Este trabajo describe el desarrollo y el uso de un modelo explicativo que genere explicaciones en lenguaje natural a partir de información de sistemas, modelos, conjuntos de datos, evaluación de modelos, entre otros aspectos relacionados con un sistema basado en Inteligencia Artificial. Esta información es gestionada a través de una ontología propuesta para Inteligencia Artificial Explicable y las explicaciones generadas se brindan a través de un chatbot conversacional.

Con base en los resultados obtenidos, se concluye que el modelo explicativo ha sido capaz de generar explicaciones textuales eficientes para las preguntas de competencia descritas. Esto ha permitido a que los sistemas de inteligencia artificial explicable desarrollados para las pruebas con usuarios sean valorados positivamente por los participantes de forma cualitativa y cuantitativa. A través de las apreciaciones de los participantes, se ha identificado que el punto de mejora más importante es la variabilidad en la forma de realizar las preguntas, el cual ha sido comentado en la sección de Discusión. Otro punto por mejorar para próximos estudios es incluir un glosario más extenso del dominio del sistema, para que el chatbot pueda facilitar el uso a los usuarios nuevos o con menor conocimiento del dominio.

En los trabajos futuros se planea abordar los siguientes temas: primero, incorporar más preguntas y respuestas tanto en número como en variabilidad a partir de una encuesta previa con usuarios; segundo, aumentar las clases y propiedades de la ontología para explicabilidad de forma que sea capaz de representar un mayor conocimiento de explicabilidad y del dominio o negocio en que se desarrolla el modelo de Inteligencia Artificial; y en tercer lugar, explorar el uso de técnicas de recuperación de información para obtener un sistema de inteligencia artificial explicable completamente automático en el cual el desarrollador del sistema no tenga que diseñar las consultas en la ontología. De esta forma, el desarrollador únicamente se dedicará a implementar el modelo de Inteligencia Artificial y a crear los individuos necesarios en la ontología, es decir, gestionar el conocimiento del sistema de inteligencia artificial explicable.

Recomendaciones

En esta sección se brindan indicaciones para facilitar el entendimiento del lector y propiciar investigaciones relacionadas. Las recomendaciones están dirigidas para dos tipos de lectores: lectores con interés en inteligencia artificial explicable y lectores con interés en procesamiento de lenguaje natural.

Para aquellos lectores que estén interesados en sistemas de inteligencia artificial explicable se sugiere prestar especial atención a las siguientes áreas:

- Preguntas de competencia: Dado que las explicaciones generadas responden estas preguntas, se recomienda medir la eficacia de estas respuestas en diferentes grupos de usuarios. Esto permitirá evaluar la robustez del sistema y su capacidad para manejar consultas de mayor complejidad.
- Estructura de la Ontología XAIO: Se recomienda profundizar en la estructura de XAIO, de manera que se puedan considerar situaciones de explicabilidad adicionales y permitir la creación de sistemas más completos y efectivos de inteligencia artificial explicable.

Si se trata de lectores con interés en procesamiento de lenguaje natural, y específicamente en generación de lenguaje natural, se recomienda enfatizar en los siguientes puntos:

- Variabilidad en las respuestas: Siendo una de las áreas de mejora identificadas en este trabajo se sugiere explorar técnicas de aumento de datos y diversidad en la generación de texto. Esto podrá enriquecer las respuestas del chatbot, haciendo que sean más naturales y, por lo tanto, confiables.
- Evaluación del modelo de generación de lenguaje natural: Para mejorar este modelo es importante realizar la evaluación con diferentes métricas para situaciones específicas dependiendo del objetivo como la fluidez o la coherencia.

Referencias Bibliográficas

- [1] E. B. J. E. K. L. T. L. J. M. H. N. J. C. N. V. P. Y. S. R. W. J. C. Nestor Maslej, Loredana Fattorini and R. Perrault, "The AI index 2023 Annual Report," *Tech. Rep., AI Index Steering Committee, Institute for Human-Centered AI, Stanford University*, 2023.
- [2] G. M. Campedelli, "Where are we? using SCOPUS to map the literature at the intersection between artificial intelligence and research on crime," *Journal of Computational Social Science*, vol. 4, no. 2, pp. 503–530, 2021.
- [3] B. H. Van der Velden, H. J. Kuijf, K. G. Gilhuijs, and M. A. Viergever, "Explainable artificial intelligence (XAI) in deep learning-based medical image analysis," *Medical Image Analysis*, p. 102470, 2022.
- [4] H. Khosravi, S. B. Shum, G. Chen, C. Conati, Y.-S. Tsai, J. Kay, S. Knight, R. Martinez-Maldonado, S. Sadiq, and D. Gašević, "Explainable artificial intelligence in education," *Computers and Education: Artificial Intelligence*, vol. 3, p. 100074, 2022.
- [5] M. R. B. Mohsin, S. A. Ramisa, M. Saad, S. H. Rabbani, S. Tamkin, F. B. Ashraf, and M. T. Reza, "Classifying insect pests from image data using deep learning," in *2022 15th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pp. 1–6, IEEE, 2022.
- [6] R. Alhomsy and A. S. Vivacqua, "The explainable business process (XBP) - an exploratory research," *Revista de Informática Teórica e Aplicada*, vol. 28, no. 1, pp. 78–96, 2021.
- [7] C. A. Zhang, S. Cho, and M. Vasarhelyi, "Explainable artificial intelligence (XAI) in auditing," *International Journal of Accounting Information Systems*, vol. 46, p. 100572, 2022.
- [8] R. Machlev, L. Heistrene, M. Perl, K. Levy, J. Belikov, S. Mannor, and Y. Levron, "Explainable artificial intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities," *Energy and AI*, p. 100169, 2022.

- [9] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you?.Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144, 2016.
- [10] H. de Bruijn, M. Warnier, and M. Janssen, "The perils and pitfalls of explainable ai: Strategies for explaining algorithmic decision-making," *Government Information Quarterly*, vol. 39, no. 2, p. 101666, 2022.
- [11] T. Miller, P. Howe, and L. Sonenberg, "Explainable AI: Beware of inmates running the asylum or: How I learnt to stop worrying and love the social and behavioural sciences," *arXiv preprint arXiv:1712.00547*, 2017.
- [12] T. Lei, R. Barzilay, and T. Jaakkola, "Rationalizing neural predictions," *arXiv preprint arXiv:1606.04155*, 2016.
- [13] I. Donadello, M. Dragoni, and C. Eccher, "Persuasive explanation of reasoning inferences on dietary data," in *Joint Proceedings of the 6th International Workshop on Dataset PROFILing and Search & the 1st Workshop on Semantic Explainability co-located with the 18th International Semantic Web Conference (ISWC 2019)*, vol. 2465, pp. 46–61, CEUR-WS. org, 2019.
- [14] N. Ammar, P. Zareie, M. E. Hare, L. Rogers, S. Madubonwu, J. Yaun, and A. Shaban-Nejad, "SPACES: Explainable multimodal AI for active surveillance, diagnosis, and management of adverse childhood experiences (ACEs)," in *2021 IEEE International Conference on Big Data (Big Data)*, pp. 5843–5847, IEEE, 2021.
- [15] N. Shi, Q. Zeng, and R. Lee, "XAI language tutor—a XAI-based language learning chatbot using ontology and transfer learning techniques," *International Journal on Natural Language Computing (IJNLC)*, vol. 9, 2020.
- [16] T. Tang, J. Li, W. X. Zhao, and J.-R. Wen, "MVP: Multi-task supervised pre-training for natural language generation," *arXiv preprint arXiv:2206.12131*, 2022.
- [17] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "BLEU: a method for automatic evaluation of machine translation," in *Proceedings of the 40th*

- annual meeting of the Association for Computational Linguistics*, pp. 311–318, 2002.
- [18] I. Papastratis, “Introduction to explainable artificial intelligence (XAI),” <https://theaisummer.com/>, 2021.
- [19] B. Galitsky and B. Galitsky, “Conversational explainability,” *Artificial Intelligence for Customer Relationship Management: Keeping Customers Informed*, pp. 415–445, 2020.
- [20] Y. Sun and S. S. Sundar, “Exploring the effects of interactive dialogue in improving user control for explainable online symptom checkers,” in *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pp. 1–7, 2022.
- [21] I. Celino, “Who is this explanation for? Human intelligence and knowledge graphs for explainable AI,” in *Knowledge Graphs for eXplainable Artificial Intelligence: Foundations, Applications and Challenges*, pp. 276–285, IOS Press, 2020.
- [22] A. Basu and R. Ahad, “Using a relational database to support explanation in a knowledge-based system,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 4, no. 6, pp. 572–581, 1992.
- [23] P. Inoue Nakagawa, “Semantic description of explainable machine learning workflows,” *Master’s thesis, University of Twente*, 2021.
- [24] D. Gunning and D. Aha, “DARPA’s explainable artificial intelligence (XAI) program,” *AI magazine*, vol. 40, no. 2, pp. 44–58, 2019.
- [25] A. Khurana, P. Alamzadeh, and P. K. Chilana, “ChatrEx: designing explainable chatbot interfaces for enhancing usefulness, transparency, and trust,” in *2021 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pp. 1–11, IEEE, 2021.
- [26] V. B. Nguyen, J. Schlötterer, and C. Seifert, “Explaining machine learning models in natural conversations: Towards a conversational XAI agent,” *arXiv preprint arXiv:2209.02552*, 2022.
- [27] H. Fanaee-T and J. Gama, “Event labeling combining ensemble detectors and background knowledge,” *Progress in Artificial Intelligence*, vol. 2, pp. 113–127, 2014.
- [28] D. Ganatara, “Campus Recruitment Dataset,” 2019.

- [29] UCI Machine Learning, “Pima Indians Diabetes Database,” 2016.
- [30] D. Aha and D. Kibler, “Instance-based prediction of heart-disease presence with the Cleveland database,” *University of California*, vol. 3, no. 1, pp. 3–2, 1988.
- [31] L. Breiman, “Random forests,” *Machine learning*, vol. 45, pp. 5–32, 2001.
- [32] D. Esteves, A. Ławrynowicz, P. Panov, L. Soldatova, T. Soru, J. Vanschoreng, “ML Schema CG Specification,” 2016.
- [33] S. Chari, O. Seneviratne, D. M. Gruen, M. A. Foreman, A. K. Das, and D. L. McGuinness, “Explanation ontology: a model of explanations for user-centered AI,” in *The Semantic Web–ISWC 2020: 19th International Semantic Web Conference, Athens, Greece, November 2–6, 2020, Proceedings, Part II*, pp. 228–243, Springer, 2020.
- [34] Q. Chen, Z. Zhuo, and W. Wang, “BERT for joint intent classification and slot filling,” *arXiv preprint arXiv:1902.10909*, 2019.
- [35] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” *arXiv preprint arXiv:1810.04805*, 2018.

