ISBN: 978-628-95207-8-1. ISSN: 2414-6390. Digital Object Identifier: 10.18687/LACCEI2024.1.1.1588

ChatGPT in the Generation of Explanations for Cyber-Physical Systems

Oscar Peña-Cáceres, Dr[1,2], Henry Silva-Marchan, Mg[3], Rudy Espinoza-Nima, Mg[4], Elvis Garay-Silupu, Ing[2], Dania Ricalde-Morán, Mg[3], and Douglas Alvarado-Paiva, Dr[4]
[1]Departament d'Informàtica, Universitat de València, Spain, osjmarpe@alumni.uv.es
[2]Universidad César Vallejo, Peru, ojpenac@ucvvirtual.edu.pe; egarays@ucvvirtual.edu.pe
[3]Universidad Nacional de Tumbes, Peru, hsilvam@untumbes.edu.pe; dmricaldem@untumbes.edu.pe
[4]Universidad Nacional de Piura, Peru, respinozan@unp.edu.pe; dalvaradop@unp.edu.pe

Abstract*– The research focused on designing a mechanism for the generation of explanations in the smart home environment through ChatGPT with the purpose of dynamizing the interaction between humans and cyber-physical systems. A simulator was developed that integrates different services and explanation objects related to the smart home environment. The simulator interacts with an artificial intelligence model to determine the level of attention required by the user, the type of explanation, and the interaction mechanism. The ChatGPT text-davinci-003 model was used to generate explanations. For the evaluation, a checklist was applied to 20 professionals linked to the study area in order to determine the quality of the system and the effectiveness of ChatGPT in the generation of explanations for cyber-physical systems in the context of the smart home. The system provided favorable results, with outstanding scores on the quality of explanations (95%), user-system interaction (90%), and adaptability and robustness (95%). The explanations varied between 15 and 50 words, depending on the scenario. Each explanation generated by the system has proven to be clear and understood by users. The explanations that focus on answering the "What" of a scenario are brief and the "Why" explanations are longer. For future research, it is recommended to explore the GPT-3.5-turbo-instructions and Davinci-002 models, as well as to consider the inclusion of new forms of explanation, referring to "why not" and "how" the user should execute the tasks that the system has inferred.

***Keywords*-- Explanations, ChatGPT, Systems, Cyber-Physical.**

Digital Object Identifier: (only for full papers, inserted by LACCEI).
ISSN, ISBN: (to be inserted by LACCEI).
DO NOT REMOVE

ChatGPT en la Generación de Explicaciones para Sistemas Cyber-Físicos

Oscar Peña-Cáceres, Dr[1,2], Henry Silva-Marchan, Mg[3], Rudy Espinoza-Nima, Mg[4], Elvis Garay-Silupu, Ing[2], Dania Ricalde-Morán, Mg[3] y Douglas Alvarado-Paiva, Dr[4]
[1]Departament d'Informàtica, Universitat de València, España, osjmarpe@alumni.uv.es
[2]Universidad César Vallejo, Perú, ojpenac@ucvvirtual.edu.pe; egarays@ucvvirtual.edu.pe
[3]Universidad Nacional de Tumbes, Perú, hsilvam@untumbes.edu.pe; dmricaldem@untumbes.edu.pe
[4]Universidad Nacional de Piura, Perú, respinozan@unp.edu.pe; dalvaradop@unp.edu.pe

Abstract*– La investigación se enfocó en diseñar un mecanismo para la generación de explicaciones en el ámbito del hogar inteligente mediante ChatGPT con el propósito de dinamizar la interacción entre humanos y sistemas cyber-físicos. El simulador dispone de una variedad de servicios, como gestión de riego, residuos, entre otros, junto con objetos explicativos que están vinculados al lugar u objeto con el cual el usuario interactúa. Las entradas están diseñadas para identificar el nivel de atención necesario por parte del residente, el tipo de explicación requerido y el mecanismo de interacción, todo basado en un modelo pre-entrenado. Se utilizo el modelo text-davinci-003 de ChatGPT 3.5 para la generación de explicaciones. Para la evaluación, se aplicó una lista de verificación dirigida a 20 profesionales vinculados al área de estudio. La iniciativa proporcionó resultados favorables, con puntuaciones destacadas sobre la calidad de explicaciones (95%), interacción usuario-sistema (90%), y adaptabilidad y robustez (95%). Las explicaciones, varían entre 15 y 50 palabras según el escenario. Cada explicación generada por el sistema ha demostrado ser clara y comprendida por los usuarios. Las explicaciones que se inclinan en responder el "Qué" de un escenario son breves y las del tipo "Porqué" más extensas. Para futuras investigaciones, se recomienda explorar los modelos GPT-3.5-turbo-instrucciones y davinci-002, además de considerar la inclusión de nuevas formas de explicación, referidas al "Por qué no" y "Cómo" el usuario debería ejecutar las tareas que el sistema ha inferido.

Keywords—Explicaciones, ChatGPT, Sistemas, Cyber-Físicos.

I. Introducción

La Inteligencia Artificial (IA) y el Aprendizaje Automático (ML) han experimentado avances importantes en los últimos tiempos, y ahora se utilizan ampliamente en una variedad de campos para crear sistemas automatizados o semiautomatizados. La aceptación de estos sistemas por parte de la sociedad ha generado un notable avance en el campo de la IA. No obstante, a pesar de la disponibilidad de modelos altamente precisos, la falta de explicabilidad e interpretabilidad constituye un desafío importante que se debe abordar. Un problema adicional es que, en ciertas ocasiones, estos sistemas pueden inundar al usuario con una cantidad excesiva de información, lo que podría generar confusión y dificultar la toma de decisiones.

Complementando lo manifestado, Ronanki et al. [7] señala que existe una carencia de métodos y metodologías flexibles que se pongan en práctica el desarrollo de este tipo de sistemas. Esta ausencia de modelos ha resultado que los robots o sistemas de monitorización carezcan de un componente que proporcione explicaciones comprensibles para los humanos sobre las acciones o tareas ejecutadas por el sistema [8]. A pesar que se crea que lo sistemas autónomos podrían sustituir al ser humano, estos tienen la intención de colaborar en los procesos por los cuales fueron diseñados [6].

Esta brecha representa una oportunidad para abordar investigaciones que se centren en desarrollar métodos específicos dirigidos a diferentes dominios y tareas [1], donde los usuarios puedan comprender la relación entre la acción decidida por el sistema y la participación humana requerida. En ese sentido, creemos que es viable diseñar sistemas que ofrezcan explicaciones claras y comprensibles sobre las acciones ejecutadas. Por ejemplo, en el ámbito del hogar inteligente, entender por qué se incrementa la temperatura ambiente o se encienden las luces del portal. Esta práctica no solo podría consolidar la confianza del usuario, sino también mejorar la interacción y la participación de los residentes.

Para este caso, Chen et al. [2] propuso una estructura narrativa y causal en el escenario de la consulta sanitaria. En paralelo [3], precisa que los enfoques actuales para generar explicaciones basadas en frases tienen algunas limitaciones, lo que restringe la expresividad de las frases, por lo que sugieren optar por la generación de frases de estilo libre, lo que podría mejorar la calidad de las explicaciones. También, descubrimos que un sistema que integra IA podría ofrecer un nivel de eficiencia, precisión y sincronización entre el usuario y la máquina, generando así un alto nivel de confianza en las explicaciones [4]. Complementando Waa et al. [5] indica que las explicaciones contrastivas basadas en reglas y ejemplos son dos estilos de explicación ejemplares.

Por otro lado, el despliegue de ChatGPT, Google Bard, y Microsoft Bing, centradas generación de textos e imágenes [9], ha promovido el desarrollo de estudios transversales que utilizan este tipo de IA para recibir la explicación de temas sencillos o complejos, como descubrir en términos conceptuales el espacio exterior, culturas, avances tecnológicos, entre otros [10], con un lenguaje sencillo para que un niño los entienda.

En esta misma dirección el autor [9] realizó un análisis comparativo de las plataformas antes mencionadas, las variables que contemplo la investigación fue precisión, tiempo de respuesta, relevancia, satisfacción del usuario y

participación del usuario. Los resultados del estudio indican que ChatGPT superó a otras tecnologías de chatbot en términos de precisión y relevancia, mientras que Google BARD tuvo el tiempo de respuesta más rápido. Microsoft Bing demostró la mayor satisfacción y compromiso del usuario.

En este sentido, consideramos a ChatGPT como un modelo de inteligencia artificial altamente sofisticado que ha experimentado un aumento importante en su popularidad. Este modelo tiene la capacidad de comprender y generar texto en lenguaje humano y se emplea en una amplia variedad de aplicaciones, que incluyen sistemas de atención al cliente automatizados, chatbots y la creación de contenido [11]. También tiene muchos beneficios para profesores y estudiantes, especialmente en tareas basadas en texto que de forma manual pueden demandar tiempos excesivos [12]. Algunas perspectivas [13] indican que los usuarios deben utilizar ChatGPT teniendo en cuenta características como, encontrar formas de usarlos de manera efectiva, que su uso no constituya plagio [14], saber cuantificar su sesgo, que los usuarios tengan cuidado con su escasa precisión, y que su empleabilidad sea un eslabón a la investigación y una herramienta académica.

Este estudio se adentra en la utilización del servicio de ChatGPT como una herramienta para la generación de explicaciones destinadas a sistemas cyber-físicos, centrándose específicamente en el ámbito de las casas inteligentes como caso de estudio. Nuestra aspiración es que este artículo suscite un sólido interés y promueva un debate más profundo sobre la integración de servicios como ChatGPT en la generación de explicaciones.

II. Antecedentes teóricos

En este apartado se abordan los fundamentos teóricos de la Inteligencia Artificial Explicable (XAI) desde un enfoque técnico, omitiendo la filosofía y la taxonomía. La explicabilidad requiere de una interfaz que debe ser comprensible de manera simultánea para los seres humanos y proporcionar una representación precisa al responsable de la toma de decisiones [15]. En este dominio, la explicabilidad simboliza la conexión entre los modelos y los usuarios finales, permitiendo que estos últimos obtengan aclaraciones sobre las decisiones del modelo de IA/ML. Esta sección discute los aspectos necesarios para que XAI sea efectiva y confiable en diversas aplicaciones, resumiendo los conceptos clave basados en investigaciones previas [3,4].

2.1. Alcance de la Explicabilidad

Dos análisis bibliográficos recientes sobre la XAI han identificado que existen dos alcances de explicabilidad, denominados global y local [3,4]. El enfoque global, busca que todo el proceso de inferencia de un modelo sea transparente y comprensible para el usuario, como en el caso de un árbol de decisión. Por otro lado, la explicación con un alcance local se refiere a proporcionar una explicación explícita para una sola instancia de inferencia, como podría ser una rama individual en el contexto de los árboles de decisión.

2.2. Momentos de la Explicabilidad

Para Vilone y Longo, existen dos momentos clave en los que un modelo genera una explicación de la decisión [3,4]:

A. Antes de: Este enfoque consideran la generación de explicaciones desde el inicio del entrenamiento de los datos para lograr un rendimiento óptimo.

B. Después de: En esta categoría, se emplea un modelo externo o sustituto junto con el modelo base. El modelo base permanece sin cambios, mientras que el modelo externo replica su comportamiento para proporcionar explicaciones a los usuarios. Estos métodos son particularmente útiles cuando el mecanismo de inferencia del modelo base es desconocido para los usuarios, como en las máquinas de soporte vectorial y las redes neuronales.

2.3. Propiedades de la Explicabilidad

Los métodos disponibles para añadir explicabilidad a los modelos AI/ML se agrupan inicialmente en función de tres propiedades: (1) la fase de generación de una explicación; (2) el alcance de la explicación; y (3) la forma de la explicación [1].

III. Trabajos Relacionados

En los últimos dos años, se ha observado un considerable incremento en la actividad investigativa relacionada con el desarrollo de teorías, metodologías y recursos pertinentes a la XAI. Este crecimiento refleja un interés en comprender cómo los sistemas de IA toman decisiones y cómo estas decisiones pueden ser explicadas y comprendidas por los usuarios. La primera revisión documentada en la literatura sobre este tema se atribuye a la obra de Lacave y Diéz, que marcó un hito en el campo al abordar los fundamentos y los primeros enfoques para lograr la XAI [18].

La investigación de Ribeiro y colegas, se centró en realizar una revisión de modelos interpretables como una solución para abordar el desafío de dotar de explicabilidad a los modelos de Inteligencia Artificial/Aprendizaje Automático, tales como los modelos aditivos, los árboles de decisión, las redes neuronales y los modelos lineales dispersos [19]. Posteriormente, presentaron una técnica de carácter agnóstico con respecto a los modelos, la cual involucra el desarrollo conjunto de un modelo interpretable a partir de las predicciones generadas por modelos de caja negra. También el estudio de [20] sostiene que esta variedad de categorías de explicación será de utilidad para los futuros arquitectos de sistemas al momento de crear y dar prioridad a los requisitos,

así como para generar explicaciones que se adapten de manera óptima a los requisitos de los usuarios y las circunstancias.

En [21] elaboraron un modelo para producir resúmenes utilizando un modelo de lenguaje transformador. En su propuesta utilizan un modelo de nombre Pegasus el cual se caracteriza por ser un algoritmo que representa criterios lógicos y que parte de su funcionamiento se encuentra en la clasificación de términos o palabras claves que permitan la construcción de un resumen de acuerdo a los datos de entrada. Los resultados de este estudio describen que el modelo coincidía exactamente con las pruebas de entrada en el 60% de los casos. Este escenario demuestra que a medida que avanzamos hacia la inteligencia artificial, está claro que es posible dotar a los agentes como robots o sistemas inteligentes de facultades que garanticen que son dignos de confianza. En concretamente, los agentes deben ser capaces de explicarse a sí mismos de un modo que sea a la vez lógicamente correcto y comprensible para los humanos.

Hasani at. [22] indica que los modelos de aprendizaje automático han experimentado una amplia adopción en los últimos años y que la creación de explicaciones concisas y precisas tiende a elevar la confianza del usuario y mejorar la comprensión de las predicciones del modelo. Generalmente, los algoritmos de explicación más conocidos están altamente optimizados para generar explicaciones de manera individual para una única predicción. Sin embargo, en la práctica, suele ser necesario generar explicaciones en lotes para múltiples predicciones simultáneamente. Los autores precisan que, a la fecha, no se ha llevado a cabo ningún trabajo que aborde eficientemente la generación de explicaciones para más de una predicción al mismo tiempo. Aunque es posible utilizar múltiples máquinas para generar explicaciones en paralelo. En este sentido, las diversas corrientes de pensamiento acerca de lo que constituye una explicación y los autores [23] indican que el aprendizaje automático podría beneficiarse desde una perspectiva más holística sobre diferentes contextos en la que un usuario pueda recibir e interpretar una explicación proporcionada por el sistema.

Los autores [24] sostienen que la calidad del mecanismo de generación de explicaciones de un agente se basa en lo bien que cumple tres objetivos o propósitos de la producción de explicaciones que descubran patrones desconocidos u ocultos, resaltar o identificar cadenas causales relevantes e identificar suposiciones de fondo incorrectas. También presentan un sistema autoexplicativo de nombre AERA (Autocatlytic Endogenous Reflective Architecture), capaz de autoexplicarse dirigido por objetivos: Explicar de forma autónoma su propio comportamiento, así como sus conocimientos adquiridos sobre las tareas y el entorno. Entre sus reflexiones precisan que, el objetivo de crear sistemas con inteligencia artificial general, radica en que los sistemas no sólo deberían ser explicables, sino que deberían poder explicarse a sí mismos a sus usuarios.

El estudio [25] aborda el desarrollo de un chatbot denominado (EQRbot) para generar explicaciones sobre los consejos de tratamiento de los pacientes en el área de la salud. El esquema EQR bosqueja un patrón de interacciones Explicación-Pregunta-Respuesta dirigido agentes de este dominio. El funcionamiento de la solución se realiza mediante el uso de plantillas que alimentan al agente conversacional que transmitirá exhaustivamente la información solicitada y las respuestas a las consultas de los usuarios siguientes en forma de mensajes. Por otro lado, los autores enfatizan que el PNL es el principal medio para mejorar la concordancia de palabras entre la entrada del usuario y las explicaciones almacenadas en el sistema.

La investigación [8] propone una arquitectura de generación de comportamientos explicables para la interacción entre humanos y robots sociales autoexplicables de tal forma que sea generalmente interpretable y, por tanto, explicable a nivel sociocomportamental. El modelo sugerido para el flujo de diálogos explicativos en la interacción aborda peticiones donde el usuario puede preguntar ¿Qué? y ¿Por qué? el agente realiza dicha acción o cual es la razón que conllevo a que brindará algún tipo de información. Las herramientas y métodos utilizados en esta solución han sido Google ASR para el reconocimiento automático del habla, RasaNLU para la comprensión del lenguaje natural y Spacy para la clasificación de patrones. Este tipo de acercamientos podría aumentar la naturalidad de la interacción social al equilibrar más armónicamente las necesidades de sistemas inteligentes y su interacción con el usuario donde los enfoques de aprendizaje deben incorporarse de manera que coadyuven a madurar el razonamiento del agente.

IV. Planteamiento del Problema y Visión General

Los sistemas autónomos poseen la capacidad de ajustar sus acciones en respuesta a modificaciones en su entorno, llevando a cabo adaptaciones de manera autónoma. Sin embargo, este proceder autónomo puede generar confusiones entre los usuarios, quienes podrían no comprender completamente las razones detrás de los cambios en el comportamiento del sistema, generando respuestas y cooperaciones incorrectas por parte de los usuarios. Para abordar este problema de falta de comprensión y mejorar la colaboración efectiva entre humanos y sistemas autónomos, resulta esencial emplear explicaciones que intenten argumentar y comunicar a los usuarios los motivos subyacentes a las alteraciones en el comportamiento del sistema.

Estas explicaciones deben ser consideradas como elementos dinámicos, capaces de ajustarse a diversos factores internos y externos, como la capacidad, experiencia, género o estado de atención de las personas, así como a la presencia de fallos o conflictos en el sistema y las condiciones del entorno, entre otros. Es importante reconocer que la entrega de explicaciones no debería ser constante, ya que proporcionar información detallada en todo momento podría resultar abrumador para los usuarios, especialmente cuando no es necesario. Por ende, la provisión de explicaciones debería activarse únicamente cuando se identifique su necesidad. Tomemos como ejemplo una tarea en un hogar inteligente

donde el sistema realiza pedidos de compras de manera autónoma. Si, por alguna razón, el servicio al que se envía el pedido inicialmente no está disponible, el sistema se adapta redirigiendo el pedido a otro proveedor. Este cambio en el comportamiento podría desconcertar a un usuario con poca experiencia en el sistema, llevándolo a perder el entendimiento de la situación. Sin embargo, si se detecta que el usuario no comprende lo que está ocurriendo, por ejemplo, si intenta realizar la compra manualmente, en ese momento se debería ofrecer una explicación para aclarar la situación.

En este trabajo se propone una solución para la generación de explicaciones que utiliza un modelo predictivo de comprensibilidad, para inferir cuando es necesario ofrecer una explicación al usuario ante una acción de adaptación del sistema en el dominio del hogar inteligente. La arquitectura de la propuesta se muestra en la Figura 1.

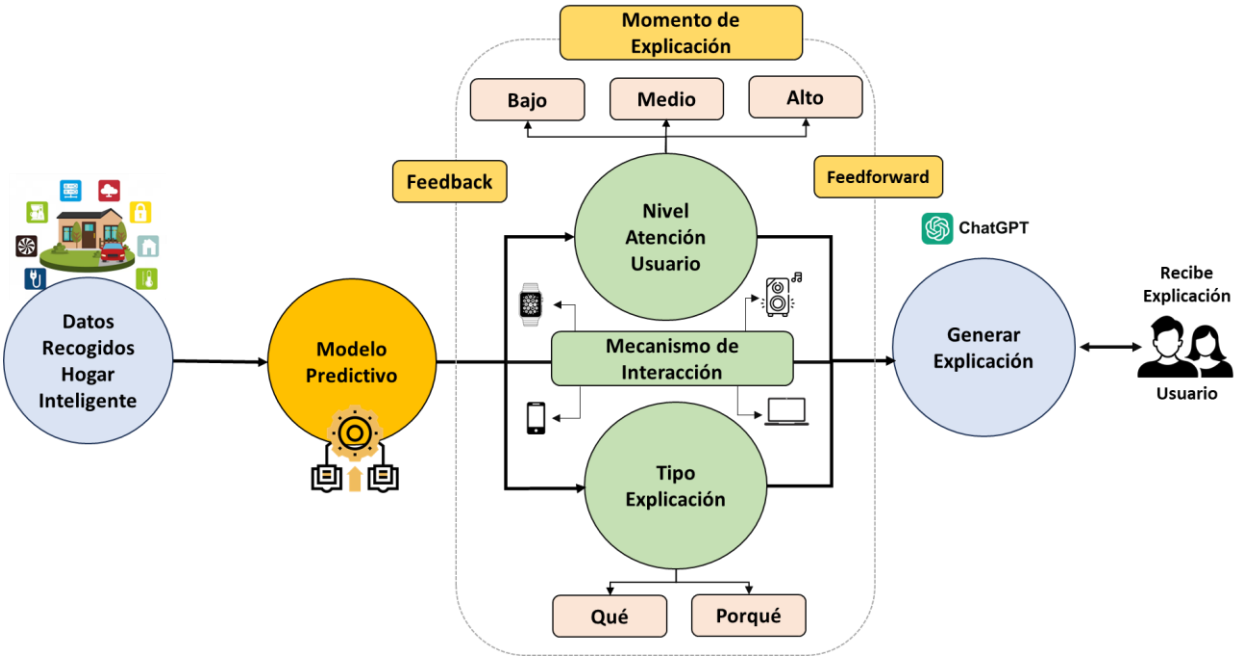

Fig. 1. Arquitectura de la solución propuesta

V. ChatGPT para la Generación de Explicaciones

En esta sección, se describe el procedimiento empleado para entrenar a ChatGPT y de qué manera es posible generar explicaciones claras.

5.1. Recuperación y limpieza de datos

Se contó con un conjunto de 354 registros de datos relacionados con dispositivos conectados en el ámbito del hogar inteligente, donde los usuarios interactuaban con el sistema mediante solicitudes como encender luces, aumentar la temperatura o cerrar persianas, entre otras acciones. Se llevó a cabo un proceso de curación y ajuste de estos datos con el fin de garantizar su relevancia y la claridad de las interacciones registradas.

5.2. Pre-entrenamiento

En esta fase, se suministraron a ChatGPT 253 de los 354 registros disponibles. Posterior a ello, se le instruyó con 15 escenarios simulados para enseñar al modelo a gestionar situaciones complejas y responder de manera efectiva a las necesidades del usuario en un ambiente conectado.

5.3. Instrucciones

En esta etapa, nos centramos en la elaboración de las instrucciones o prompts donde involucramos cuatro aspectos clave: contexto, tarea, instrucción y claridad.

- El contexto debe ser específico y detallado, proporcionando al modelo la información necesaria para entender el entorno en el que debe operar.
- La tarea debe estar delineada, con el objetivo que se desea alcanzar.
- Las instrucciones deben ser precisas y directas, evitando ambigüedades.
- En cuanto a la claridad se debe utilizar un vocabulario adecuado al dominio y al contexto proporcionado.

En esta dirección la Figura 2 ilustra cómo, mediante estos criterios, es posible generar solicitudes con fines de entrenamiento y obtener una explicación sobre un contexto especifico.

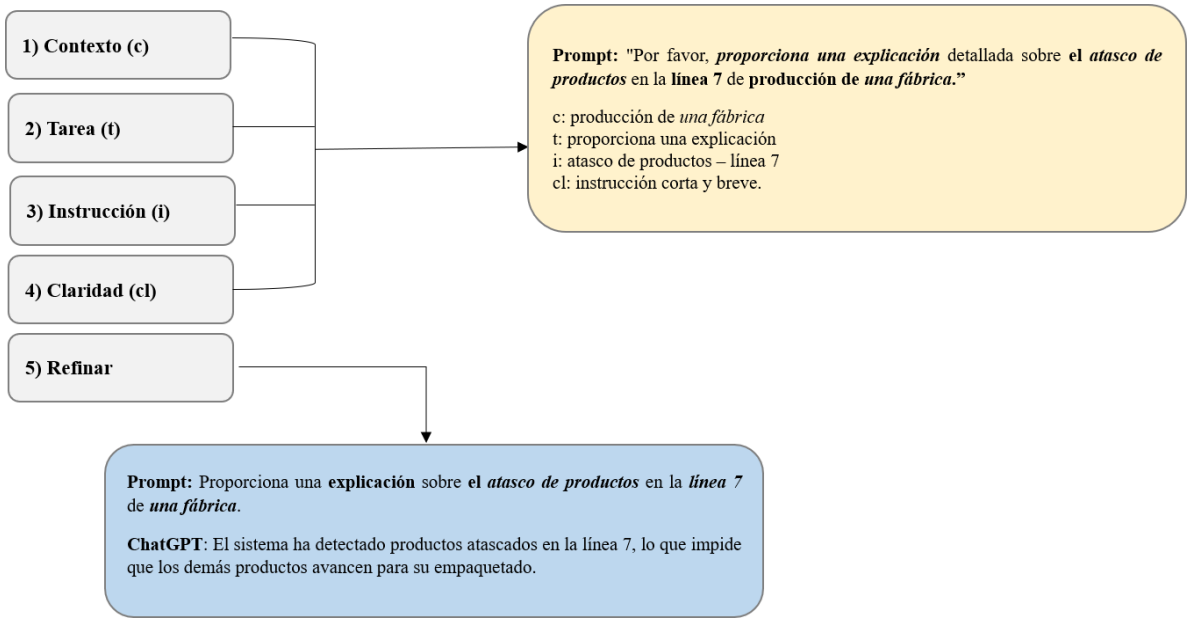

Fig. 2. Criterios para la generación de explicaciones

Al definir y caracterizar adecuadamente cada una de los patrones es posible obtener explicaciones a través de ChatGPT que se integren de manera efectiva en un dominio específico y mejoren la experiencia e interacción entre humanos y sistemas en diversos contextos.

VI. Explicaciones en el Ámbito del Hogar Inteligente

La generación de explicaciones en el ámbito del hogar inteligente se experimenta a través del servicio de ChatGPT. Al aprovechar la información recopilada de dispositivos y sistemas interconectados en un hogar inteligente, se abre la posibilidad de proporcionar respuestas y aclaraciones prácticas sobre el funcionamiento de dispositivos electrónicos específicos o tareas programadas en el entorno doméstico.

En este escenario, se plantea la viabilidad de automatizar explicaciones, abarcando desde tareas periódicas hasta la gestión de recomendaciones, mediante la aplicación de ChatGPT. Para lograr una generación efectiva de explicaciones, resulta esencial implementar solicitudes y consultas parametrizadas que comprendan el contexto y las necesidades inferidas por el sistema. La propuesta incluye la adopción de un conjunto de directrices y formatos estandarizados para consultas, facilitando así la interacción y asegurando respuestas coherentes y útiles.

Esta propuesta en un futuro no solo agilizaría la interacción entre los usuarios y sistemas autónomos, sino que también inauguraría un nuevo paradigma en la aplicación de ChatGPT, permitiendo un uso más auténtico y eficiente de la tecnología en la vida diaria.

6.1 Servicios y objetos de explicación

En la presente sección, se lleva a cabo una definición y caracterización preliminar de los servicios que intervienen en un entorno de hogar inteligente, como se detalla en la Tabla 1. Estos servicios encapsulan las funciones y habilidades necesarias para la automatización en un hogar conectado. Al identificar y delinear de manera clara estos servicios, se establece una base robusta que facilita el análisis y comprensión de cómo las explicaciones podrían integrarse en uno o más contextos para mejorar la experiencia de los residentes en un hogar inteligente.

TABLA I
Servicios en el Dominio Hogar Inteligente

Servicios	Descripción
Control de Iluminación Externa	Permite a los residentes del hogar inteligente gestionar y controlar la iluminación en áreas exteriores, como jardines, patios y entradas, mediante la automatización y el ajuste de la intensidad luminosa según las necesidades y preferencias.
Control de Iluminación Interna	Los ocupantes pueden gestionar y controlar la iluminación en el interior de la vivienda, incluyendo habitaciones, salas de estar y áreas de trabajo, optimizando la eficiencia energética y la comodidad.
Control de Cortinas y Persianas	Permite la apertura y cierre automatizado de cortinas y persianas de las ventanas, brindando control sobre la entrada de luz natural y la privacidad en diferentes momentos del día.
Control de Acceso y Cierre	Facilita la gestión segura del acceso a la vivienda mediante cerraduras inteligentes, códigos de acceso y sistemas de control de puertas, lo que aumenta la seguridad y la comodidad.

En la Tabla 2 se proporciona una descripción de los servicios y objetos-lugares de explicación asociados al entorno del hogar inteligente. Estos abarcan desde el control de la iluminación, tanto interna como externa, hasta la gestión de electrodomésticos y la seguridad del hogar. Cada servicio se encuentra enlazado a un objeto de explicación que detalla las áreas y elementos presentes en el entorno residencial. Estos elementos ofrecen una perspectiva sobre cómo generar explicaciones coherentes, permitiendo al usuario obtener una comprensión mejorada tanto antes como después de interactuar con los servicios y objetos de explicación.

TABLA II
OBJETOS DE EXPLICACIÓN EN EL DOMINIO HOGAR INTELIGENTE

Servicios	Objetos y lugares de explicación
Control de Iluminación Externa	Área de piscina a jacuzzi, Área de juegos o recreación, Caminos y pasillos exteriores, Fachada de la vivienda, Patío trasero, Jardines y Áreas de paisajismo.
Control de Iluminación Interna	Sala de estar, Comedor, Cocina, Habitación, Pasillo Oficina, Baño, Lavandería y Espacio de almacenamiento.
Control de Cortinas y Persianas	Persiana de Sala de estar, Persiana de Habitación, Persiana del Pasillo, Persiana del Baño, Persianas de la Oficina, Cortina de Sala de estar, Cortina del Comedor, Cortina de la Habitación.
Control de Acceso y Cierre	Puerta Principal, Habitación, Cocina, Baño, Espacio de almacenamiento y Garaje.

Al comprender la interconexión entre los servicios y los objetos de explicación, se abre la oportunidad de diseñar sistemas autónomos auto-explicativos y centrados en el usuario. La utilización de ChatGPT para la generación de explicaciones en tiempo real emerge como un componente relevante en la industria tecnológica. Este recurso no solo potenciaría la capacidad de los sistemas para comunicar información de manera efectiva, sino que también se presentaría como una fuente constante de innovación. La integración de ChatGPT en el ámbito tecnológico tiene el potencial de mejorar la eficiencia de la comunicación y desencadenar un flujo continuo de conocimientos, contribuyendo así a un avance más rápido en el panorama tecnológico.

6.2 Instrucciones a ChatGPT para generar explicaciones en el hogar inteligente

Aunque existen diversos escenarios o contextos en los que un usuario interactúa en el ámbito de un hogar inteligente, es esencial considerar que el tipo de explicación puede clasificarse como “feedback”, que explica una acción realizada por el sistema, o como “feedforward”, que implica la participación humana en el sistema y explica la acción que el humano debe llevar a cabo. Ambos tipos de explicaciones requieren un tratamiento independiente, donde las indicaciones proporcionadas a ChatGPT comprendan la naturaleza del propósito de la explicación. En este contexto, presentamos en la Tabla 3 las secciones, indicaciones y solicitudes que el sistema debería realizar a ChatGPT para la generación de explicaciones de tipo “feedback”.

TABLA III
PROMPT PARA LA GENERACIÓN DE EXPLICACIONES - FEEDBACK

Partes de la Explicación	Prompt a ChatGPT	Tipo de explicación
Qué	"La acción que ha inferido el sistema ha sido, "+***objetivo_explicacion***+" "+***objeto_interaccion***+". El servicio es, "+***servicio***+". El área o elemento es, "+***objeto_interaccion***+". Elabora una oración corta explicando lo que ha realizado el sistema."	Feedback
Porqué	“La acción que ha inferido el sistema ha sido, "+ ***objetivo_explicacion***+" "+***objeto_interaccion***+". El servicio es, "+***servicio***+". El área o elemento es, "+***objeto_interaccion***+". El motivo, razón o causa es, "+***motivo***+". Elabora una oración corta explicando lo que ha realizado el sistema.”	

La información presentada en la Tabla 4 no solo facilita la generación de explicaciones de manera centrada y comprensible, sino que también puede adaptarse a otros contextos o ámbitos laborales relacionados con la propuesta. Esto se debe a que las entradas están combinadas entre variables y textos para formar oraciones, que luego se convierten en solicitudes dirigidas al servicio de ChatGPT. Por otro lado, la Tabla 4 detalla los tipos de instrucciones para generar explicaciones en un contexto “feedforward”. De esta manera, se observa que ciertos fragmentos de contenido difieren de la Tabla 3, utilizando únicamente las variables de entrada que enriquecen la indicación y posibilitan la generación de la explicación.

TABLA IV
PROMPT PARA LA GENERACIÓN DE EXPLICACIONES - FEEDFORWARD

Partes de la Explicación	Prompt a ChatGPT	Tipo de explicación
Qué	"La acción que debe hacer el usuario es, "+ ***objetivo_explicacion*** +" "+***objeto_interaccion***+". El servicio es, "+***servicio***+". El momento es, ahora. El motivo o de acuerdo es, "+***motivo***+". Elabora una oración corta en tercera persona del singular del verbo en modo imperativo explicando una razón real."	Feedforward
Porqué	"La acción que debe hacer el usuario es, "+ ***objetivo_explicacion***+" "+***objeto_interaccion***+". El servicio es, "+***servicio***+". El momento es, ahora. El motivo o de acuerdo es, "+***motivo***+". Elabora una oración corta en tercera persona del singular del verbo en modo imperativo explicando una razón real."	

6.3 Simulador para la generación de explicaciones

Después de tener claro los escenarios, partes de explicación, tipos de explicación, servicios, objetos de explicación y peticiones. Esta sección describe el procedimiento utilizado para la generación de explicaciones utilizando el servicio de ChatGPT en su versión GPT-3.5. Inicialmente la API de ChatGPT se encontraba disponible para entornos de trabajo que interactuaban con el lenguaje de programación Python. A medida que ChatGPT ha ido escalando, se apertura la posibilidad de trabajar con servicios en la nube. Para este caso se utilizó el lenguaje de programación PHP y JavaScript, por su versatilidad. Gracias a estos medios se logró diseñar la interfaz-simulador[1] que se visualiza en la Figura 3. En esta representación se puede apreciar etiquetas como, el tipo de explicación (feedback o feedforward), servicio (ver Tabla 3), objeto de explicación (ver Tabla 2) y objetivo de la explicación que esta referida a la acción que el sistema ha realizado o es necesario que el usuario realice dicha acción. Recogida esta información, interactúa con un modelo de aprendizaje automático que determina que debe explicar el sistema, es decir, si la explicación debe explicar, el "Qué" y "Porque". De la misma manera el modelo, determina el mecanismo de interacción por el cual debe proporcionar la explicación, es decir, a través de un dispositivo móvil, smartwatch, tablet, portátil o altavoces. Cada una de las etiquetas y las salidas del modelo de aprendizaje automático, interactúan directamente con las peticiones descritas en la sección anterior (Tabla 3 y 4) que permiten generar las explicaciones.

Fig. 3. Interfaz de entradas y salidas para la generación de explicaciones

La participación del modelo de aprendizaje automático en esta solución de hogar inteligente permite poder realizar simulaciones más reales. Lo que significa, validar el desempeño de cada uno de las etiquetas y de esta forma conocer si es posible obtener explicaciones efectivas o que carecen de coherencia. En la Figura 5, planteamos un caso, las luces de la puerta principal de una vivienda aún no han sido encendidas a pesar de encontrarse sobre las 21:00 horas. En un estado real el sistema monitoriza en base a los sensores y configuraciones proporcionadas por el usuario y realiza el encendido de las luces de manera automática. El residente de la vivienda se preguntaría, ¿Qué acción realizó el sistema?. La explicación generada por la solución, "El sistema ha encendido la iluminación externa de la fachada de la vivienda a través del servicio de control de iluminación externa". Esta acción que ha realizado el sistema corresponde a un tipo de explicación de "feedback" en razón que ha sido ejecutada en base a los datos recolectados por el sistema de monitorización del hogar inteligente.

[1]https://mural.uv.es/osjmarpe/v2/index.html

Fig.4. Simulación para la generación de explicaciones, caso: encendido de iluminación

Por otro lado, también la Figura 4, bosqueja el mecanismo de interacción donde el residente recibe la explicación a través del móvil. Sin embargo, otra de de las cuestiones a tener en cuenta son los mensajes que remite el sistema, podrían ser recibidos como mensajes de texto o notificaciones de acuerdo a la configuración del sistema. En lo que respecta a las explicaciones del tipo "feedforward", las cuales se caracterizan por su mayor complejidad debido a la necesidad de participación activa del usuario, se destaca que en las explicaciones generadas se especifica que el usuario debe llevar a cabo la acción de encender las luces. Este enfoque busca fortalecer la participación y la confianza del usuario mediante una interacción más directa y colaborativa entre el humano y el sistema.

VII. Resultados

Para evaluar las explicaciones generadas por el simulador, se empleó un instrumento, detallado en la Tabla 4 que corresponde a una ficha de chequeo. Se contó con la participación de 20 profesionales experimentados en el área de conocimiento. Estos expertos desempeñaron un papel relevante al evaluar la calidad y coherencia de las explicaciones generadas por el sistema. La ficha de chequeo reflejó 10 puntos de verificación, proporcionando así una perspectiva valiosa para futuros trabajos. La herramienta elegida para administrar esta lista de verificación fue Google Forms. Cada experto fue debidamente informado sobre el propósito del estudio y se les concedió un plazo de 15 días para interactuar y explorar la solución de manera exhaustiva.

TABLA V
Instrumento de Evaluación

Indicadores	Preguntas
Calidad de las Explicaciones	¿Las explicaciones generadas por el simulador son claras y comprensibles para los usuarios?
	¿La información proporcionada por el simulador es precisa y relevante en relación con los escenarios planteados?
	¿Las explicaciones muestran coherencia en términos de lógica y fluidez?
	¿El simulador presenta una variedad adecuada de expresiones y términos para enriquecer las explicaciones?
Interacción Usuario-Sistema	¿La interfaz del simulador facilita la interacción del usuario para recibir y comprender las explicaciones generadas?
	¿El tiempo de respuesta del simulador es adecuado y contribuye a una experiencia de usuario eficiente?
	¿El simulador ofrece opciones de retroalimentación o aclaración cuando el usuario lo solicita?
	¿Las explicaciones generadas se adaptan a la experiencia y nivel de conocimiento del usuario?
Adaptabilidad y Robustez	¿El simulador demuestra adaptabilidad al generar explicaciones coherentes en diferentes escenarios y contextos?
	¿Las explicaciones mantienen su calidad y relevancia frente a cambios en la información de entrada?

Los resultados que se ilustran en la Figura 5, describen la evaluación del simulador para la generación de explicaciones e indican un desempeño sólido y eficiente en los diferentes aspectos analizados. La calidad de las explicaciones con un

95% de aceptación evidencia en los usuarios, explicaciones claras y coherentes en la mayoría de los casos. En cuanto a la interacción usuario-sistema el simulador obtuvo una puntuación del 90%, aunque podría mejorarse la eficiencia y la experiencia del usuario. Mientras que la adaptabilidad y robustez del simulador, recibieron una alta calificación del 95%, recalcando la capacidad del sistema para generar explicaciones en diversos escenarios y mantener su calidad ante cambios inesperados.

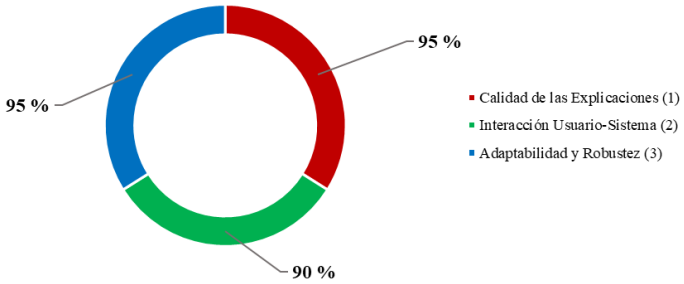

Fig.5. Resultados del cuestionario por indicador

Los resultados evidencian que el simulador ha demostrado ser altamente efectivo en la generación de explicaciones, respaldado por la opinión de 20 profesionales en el campo. Aunque las explicaciones se distinguen por su calidad y coherencia, un historial de interacciones usuario-sistema podría aún mejorar las explicaciones en el futuro.

VIII. Conclusiones

La investigación se enfocó en la generación de explicaciones a través del servicio de ChatGPT en el ámbito del hogar inteligente con la finalidad que la interacción de los sistemas cyber-físicos tengan mayor dinamismo durante la experiencia del usuario. La evaluación del sistema ha sido aceptable, tanto para la calidad de las explicaciones generadas (95%), la interacción usuario-sistema (90%) y la adaptabilidad y robustez (95%). Estos indicadores, evidencian que el sistema es competente en la generación de explicaciones de manera coherente y en una variedad de escenarios que puede extrapolarse a otros dominios, especialmente relacionados a la Industria 4.0.

El sistema genera explicaciones que oscilan entre 15 y 50 palabras, dependiendo del escenario. Las explicaciones del tipo "Qué" tienden a ser más breves, mientras que las explicaciones del tipo "Porqué" suelen ser más extensas. A pesar de la variación en la longitud, todas las explicaciones exhiben una notable facilidad de comprensión para el usuario.

Otra cuestión relevante reside en el hecho de que el sistema empleó el modelo "text-davinci-003" de ChatGPT, obteniendo explicaciones coherentes dentro del contexto. Consideramos que modelos alternativos, como "gpt-3.5-turbo-instrucciones" y "davinci-002", podrían aportar mejoras en la generación de explicaciones, utilizando términos más concisos y empáticos para el usuario.

Como trabajo futuro, se prevé la integración de nuevos tipos de explicación, que estén referidos a explicar, el "porqué no", deben omitir las indicaciones del sistema y "como" se tienen que realizar las tareas o acciones que solicita el sistema.

Agradecimientos

Agradezco al Programa Nacional de Becas y Crédito Educativo (PRONABEC) por su valioso respaldo en la etapa inicial de la investigación. Mi reconocimiento también se extiende a la Universtat de València y a mis directoras de tesis, quienes han orientado la ejecución de este primer estudio.

Referencias

[1] M. R. Islam, M. U. Ahmed, S. Barua, and S. Begum, "A Systematic Review of Explainable Artificial Intelligence in Terms of Different Application Domains and Tasks," *Appl. Sci.*, vol. 12, no. 3, Feb. 2022, doi: 10.3390/APP12031353.

[2] W.-L. Chen, A.-Z. Yen, H.-H. Huang, and H.-H. Chen, "Learning to Generate Explanation from e-Hospital Services for Medical Suggestion," pp. 2946–2951, 2022, Accessed: Sep. 12, 2023. [Online]. Available: https://www.mendeley.com/catalogue/838d1990-c2b1-364c-acb3-e4d7f27a1e6e/.

[3] L. Li, Y. Zhang, and L. Chen, "Generate Neural Template Explanations for Recommendation," *Int. Conf. Inf. Knowl. Manag. Proc.*, pp. 755–764, Oct. 2020, doi: 10.1145/3340531.3411992.

[4] C. Kim, X. Lin, C. Collins, G. W. Taylor, and M. R. Amer, "Learn, Generate, Rank, Explain: A Case Study of Visual Explanation by Generative Machine Learning," *ACM Trans. Interact. Intell. Syst.*, vol. 11, no. 3–4, Aug. 2021, doi: 10.1145/3465407.

[5] J. van der Waa, E. Nieuwburg, A. Cremers, and M. Neerincx, "Evaluating XAI: A comparison of rule-based and example-based explanations," *Artif. Intell.*, vol. 291, Feb. 2021, doi: 10.1016/J.ARTINT.2020.103404/EVALUATING_XAI_A_COMPARISON_OF_RULE_BASED_AND_EXAMPLE_BASED_EXPLANATIONS.PDF.

[6] J. Sifakis and D. Harel, "Trustworthy Autonomous System Development," *ACM Trans. Embed. Comput. Syst.*, vol. 22, no. 3, pp. 1–24, May 2023, doi: 10.1145/3545178.

[7] K. Ronanki, B. Cabrero-Daniel, J. Horkoff, and C. Berger, "RE-centric Recommendations for the Development of Trustworthy(er) Autonomous Systems," *ACM Int. Conf. Proceeding Ser.*, Jul. 2023, doi: 10.1145/3597512.3599697.

[8] S. Stange, T. Hassan, F. Schröder, J. Konkol, and S. Kopp, "Self-Explaining Social Robots: An Explainable Behavior Generation Architecture for Human-Robot Interaction," *Front. Artif. Intell.*, vol. 5, Apr. 2022, doi: 10.3389/FRAI.2022.866920/PDF.

[9] S. Bhardwaz and J. Kumar, "An Extensive Comparative Analysis of Chatbot Technologies - ChatGPT, Google BARD and Microsoft Bing," pp. 673–679, Jun. 2023, doi: 10.1109/ICAAIC56838.2023.10140214.

[10] J. Kocoń *et al.*, "ChatGPT: Jack of all trades, master of none," *Inf. Fusion*, vol. 99, Nov. 2023, doi: 10.1016/J.INFFUS.2023.101861.

[11] O. Oviedo-Trespalacios *et al.*, "The Risks of Using Chatgpt to Obtain Common Safety-Related Information and Advice," *SSRN Electron. J.*, 2023, doi: 10.2139/ssrn.4370050.

[12] M. C. Keiper, "ChatGPT in practice: Increasing event planning efficiency through artificial intelligence," *J. Hosp. Leis. Sport Tour. Educ.*, vol. 33, Nov. 2023, doi: 10.1016/J.JHLSTE.2023.100454.

[13] J. G. Meyer *et al.*, "ChatGPT and large language models in academia: opportunities and challenges," *BioData Min.*, vol. 16, no. 1, Dec. 2023, doi: 10.1186/S13040-023-00339-9.

[14] T. Fütterer *et al.*, "ChatGPT in education: global reactions to AI innovations," *Sci. Rep.*, vol. 13, no. 1, p. 15310, Sep. 2023, doi: 10.1038/S41598-023-42227-6.

[15] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A Survey of Methods for Explaining Black Box Models," *ACM Comput. Surv.*, vol. 51, no. 5, Aug. 2018, doi: 10.1145/3236009.

[16] G. Vilone and L. Longo, "Explainable Artificial Intelligence: a Systematic Review," May 2020, Accessed: Sep. 22, 2023. [Online]. Available: https://arxiv.org/abs/2006.00093v4.

[17] G. Vilone and L. Longo, "Classification of Explainable Artificial Intelligence Methods through Their Output Formats," *Mach. Learn. Knowl. Extr. 2021, Vol. 3, Pages 615-661*, vol. 3, no. 3, pp. 615–661, Aug. 2021, doi: 10.3390/MAKE3030032.

[18] C. Lacave and F. J. Díez, "A review of explanation methods for Bayesian networks," *Knowl. Eng. Rev.*, vol. 17, no. 2, pp. 107–127, Jun. 2002, doi: 10.1017/S026988890200019X.

[19] M. T. Ribeiro, S. Singh, and C. Guestrin, "Model-Agnostic Interpretability of Machine Learning," Jun. 2016, Accessed: Sep. 22, 2023. [Online]. Available: https://arxiv.org/abs/1606.05386v1.

[20] S. Chari, D. M. Gruen, O. Seneviratne, and D. L. Mcguinness, "arXiv : 2003 . 07523v1 [cs . AI] 17 Mar 2020 Directions for Explainable Knowledge-Enabled Systems," no. March, p. undefined-undefined, 2020, Accessed: Sep. 12, 2023. [Online]. Available: https://www.mendeley.com/catalogue/c69b97b7-0e21-31f2-a6fa-180aa0729ae9/.

[21] M. Giancola, S. Bringsjord, and N. S. Govindarajulu, "Toward Generating Natural-Language Explanations of Modal-Logic Proofs," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 13539 LNAI, pp. 220–230, 2023, doi: 10.1007/978-3-031-19907-3_21.

[22] S. Hasani, S. Thirumuruganathan, N. Koudas, and G. Das, "Shahin: Faster Algorithms for Generating Explanations for Multiple Predictions," *Proc. ACM SIGMOD Int. Conf. Manag. Data*, pp. 2235–2243, 2021, doi: 10.1145/3448016.3457332.

[23] B. Mittelstadt, C. Russell, and S. Wachter, "Explaining explanations in AI," *FAT* 2019 - Proc. 2019 Conf. Fairness, Accountability, Transpar.*, pp. 279–288, Jan. 2019, doi: 10.1145/3287560.3287574.

[24] K. R. Thórisson, H. Rörbeck, J. Thompson, and H. Latapie, "Explicit Goal-Driven Autonomous Self-Explanation Generation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 13921 LNCS, pp. 286–295, 2023, doi: 10.1007/978-3-031-33469-6_29.

[25] F. Castagna, A. Garton, P. McBurney, S. Parsons, I. Sassoon, and E. I. Sklar, "EQRbot: A chatbot delivering EQR argument-based explanations," *Front. Artif. Intell.*, vol. 6, 2023, doi: 10.3389/FRAI.2023.1045614.