Modelo Conceptual para Definir una Estrategia de SQA de Sistemas Generativos Impulsados por IA

Francisco Da Ponte, Camilo Pérez, Maria Pérez
fadaponte.12@est.ucab.edu.ve, ceperez.12@est.ucab.edu.ve, maperepu@ucab.edu.ve

Escuela de Ingeniería Informática, Universidad Católica Andrés Bello, Caracas, Venezuela

Resumen: Un software de calidad es aquel que cumple con los requerimientos funcionales y no funcionales, y el aseguramiento de la calidad de software depende del empleo de actividades y/o herramientas previamente definidas enfocadas en la evaluación del mismo para confirmar que cumple con sus requerimientos, como es el caso de las métricas. Los sistemas generativos impulsados por IA poseen características que hacen inefectivas muchas de las técnicas de SQA tradicional. Entre los aspectos importantes para la evaluación de calidad de sistemas basados en ML se tienen la integridad de los datos y la robustez del modelo. En el caso de los chatbots basados en sistemas generativos que emplean ML, su interacción directa y constante con los usuarios hace altamente dinámico su proceso de aprendizaje y desarrollo y crea en los usuarios la expectativa de interacciones cada vez más cercanas al trato humano, es por esto y otras características que el aseguramiento de la calidad en estos sistemas se hace muy complejo; de allí que el objetivo de este trabajo es presentar un modelo conceptual, producto de una investigación en curso, que permita desarrollar posteriormente una estrategia de evaluación de la calidad para este tipo de sistemas. El desarrollo de este modelo conceptual se logra siguiendo el Framework Metodológico Sistémico en su fase Diagnóstico, propuesto por Pérez et al.

Palabras Clave: Inteligencia Artificial (IA); Machine Learning (ML); Aseguramiento de la Calidad de Software (SQA); Pruebas; Sistemas Generativos; Chatbots.

1. Introducción

En años recientes la tecnología de inteligencia artificial (IA) ha tenido grandes avances en un corto plazo de tiempo, lo que ha llevado al desarrollo de un gran número de aplicaciones que emplean esta tecnología; sin embargo, como afirman G. Fujii et al. [1] el desarrollo de técnicas para el aseguramiento de la calidad de este tipo de software no ha avanzado al mismo paso, dejando muchas carencias en el campo de la calidad de software para estos proyectos. Por su parte, A. Carleton et al. [2] resaltan la gran necesidad que existe de desarrollar pruebas y técnicas de análisis para el software impulsado por IA; dada la falta de garantías que causa la inhabilidad de probar, verificar y certificar software impulsado por IA de manera segura y confiable. En este trabajo se presenta como resultado de una investigación en progreso, y producto del Diagnóstico del Framework Metodológico Sistémico [3], un modelo conceptual considerando aspectos tales como IA, el Machine Learning (ML) y las redes neuronales, el aseguramiento de la calidad de software (SQA) y las aplicaciones prácticas de la IA. Gracias a él se identifican las relaciones entre estos conceptos, lo cual es la base para formular una estrategia de aseguramiento de la calidad de softwares impulsados por IA del tipo sistemas generativos.

2. Inteligencia Artificial

Se comienza con los conceptos de Inteligencia e IA, según R. Sternberg [4], J. McCarthy [5] y B. Copeland [6] el término se aplica frecuentemente al desarrollo de sistemas dotados con los procesos intelectuales característicos de los seres humanos, tales como la habilidad de razonar, descubrir significados, generalizar o aprender de experiencias pasadas. Además, se observa que la IA es un área del conocimiento que pretende estudiar y desarrollar en las máquinas la inteligencia más avanzada que se conoce, que es la inteligencia humana; y que la IA no es más que conferir a un computador las cualidades que componen a esta inteligencia. También es de interés resaltar que, mediante la implementación de la inteligencia artificial, es posible estudiar y comprender aún más la inteligencia humana.

Para poder afirmar que un organismo, ya sea orgánico o artificial, es inteligente; es necesario que posea la habilidad de desarrollar la inteligencia desde el momento de su creación; por lo que un punto clave es la capacidad de aprender, ya sea a procesar información, tomar decisiones o aprender de errores o aciertos ocurridos en el pasado, es por esto que el aprendizaje es un concepto importante para comprender la inteligencia.

Los autores L. Crow y A. Crow [7], y D. Coon [8] coinciden en que el aprendizaje involucra la modificación del comportamiento con base en la experiencia; para que la IA pueda procesar la información que se le suministra y tomar decisiones basándose en ello es necesario primero que se le confiera la habilidad de aprender; por esto se han creado distintas ramas para el aprendizaje en la IA, como lo son el machine learning y el deep learning (DL).

Según T. Mitchell [9] y A. Samuel [10], ML es la capacidad de la IA de aprender por sí misma, por un plazo de tiempo indefinido, y en la que la intervención humana durante el proceso de aprendizaje, se limita a la introducción de datos.

Existen varios tipos de aprendizaje empleados en el ML. S. Russell y P. Norvig [11], y Y. Bengio et al. [12] definen estos tipos de aprendizaje, y señalan que la formulación de ellos viene dada según la necesidad particular de la aplicación en la que se pretende emplear la IA; la decisión sobre cuál tipo de aprendizaje se va a emplear viene dada por factores como los

datos disponibles y su estado particular (si se encuentran clasificados o no) al momento de alimentar a la máquina, o el nivel de precisión que se busca obtener del procesamiento de los datos por parte de la máquina.

Una forma popular de implementar estos algoritmos de aprendizaje es el DL o las redes neuronales artificiales, términos que, según se aprecia por las definiciones de S. Russell y P. Norvig [11], y Y. Bengio et al. [12] suelen ser usados indistintamente uno del otro para referirse al mismo concepto. Como resultado de esto, uno de los nombres por los que se hace pasar el DL es redes neuronales artificiales. Mientras que IBM [13] señala una distinción, acotando que el "deep" en deep learning solo hace referencia a la profundidad de las capas en una red neuronal, y clasificándolas según su número de capas, siendo las de tres capas o menos las más básicas y las más complejas partiendo de más de tres capas en adelante.

La Figura 1 muestra los conceptos sobre IA hasta ahora identificados y sus relaciones, siguiendo la notación de UML para Diagramas de clase.

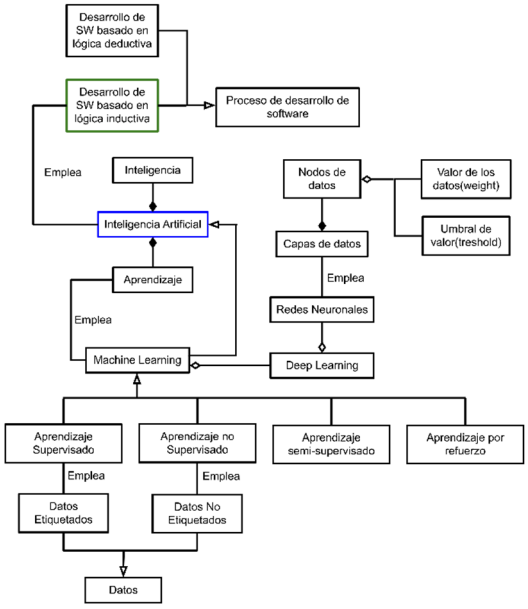

Figura 1: Modelo Conceptual de los Conceptos Expuestos sobre Inteligencia Artificial (IA). Fuente: Elaboración Propia

Ahora bien, si se revisa el proceso de desarrollo de software, se cuenta con dos vertientes, el de razonamiento deductivo o de lógica deductiva y el de razonamiento inductivo o de lógica inductiva. J. Evans et al. [14] y G. Fujii et al. [1] definen el razonamiento deductivo según su aplicación en la ciencia, afirmando que este razonamiento se basa en el conocimiento adquirido a través de la experiencia. Mientras que S-H. Nienhyus-Cheng y R. de Wolf [15], y G. Fujii et al. [1] definen el razonamiento inductivo como la empleada en el aprendizaje automático y que involucra el aprendizaje a través de ejemplos.

Para el desarrollo de software basado en ML el proceso inductivo parece ser el más eficiente, por basarse en un enfoque lógico que coincide con el proceso de aprendizaje de la máquina; en el cual se la alimenta con conjuntos de datos particulares y se le permite sacar conclusiones o dar respuesta a preguntas cada vez más complejas a medida que va aprendiendo.

3. Aseguramiento de la Calidad de SW

Como es el caso con cualquier producto, proceso o servicio derivado de la ingeniería es imprescindible evaluar el desempeño de la IA, para ello se deben definir conceptos como la calidad y su disciplina como el aseguramiento de la calidad.

De la definición de calidad de la serie de normas ISO 9000 [16] y de J. Juran y J. De Feo [17] se puede afirmar que, en esencia, un producto o servicio de calidad es aquel que posee la capacidad de cumplir con los objetivos para los cuales se ha diseñado y tiene la menor cantidad de fallas posible.

Dentro del proceso de desarrollo de software, el aseguramiento de la calidad se considera una disciplina específica denominada Software Quality Assurance (SQA). Ahora bien, para comprender el SQA es necesario definir en qué consiste la calidad de software.

La serie de normas ISO 25010 (2011) [18] y G. Fujii et al. [1] coinciden afirmando que un producto de software de alta calidad, debe funcionar bien incluso cuando algo sale mal, y que los productos de IA por su complejidad sufren de algo conocido como "CACE" o "Cambia Algo, Cambia Todo" por sus siglas en inglés.

Volviendo a SQA, se observa que R. Pressman [19] argumenta que un software de calidad es aquel que cumple con los requisitos funcionales y no funcionales establecidos por el cliente, y en que su aseguramiento (SQA) depende del empleo de un conjunto de actividades previamente definidas enfocadas en la evaluación del software y la recopilación de evidencia con el fin de confirmar que el software cumple con los requerimientos necesarios, esto a su vez, con el propósito de generar confianza en los stakeholders.

Un concepto esencial en SQA es el de las métricas, R. Pressman [19] y M. Estayno et al. [20] las definen como una herramienta de medición que permite cuantificar la calidad de un producto de software, dándole un significado práctico a la evaluación de calidad en torno a medidas precisas, compuestas de un método de medición y una escala.

G. Fujii et al. [1] y A. Carleton et al. [2] coinciden en que, el software potenciado por ML posee características particulares que en comparación con el software tradicional hace inefectivas a muchas de las técnicas empleadas en el SQA tradicional, es importante el desarrollo de enfoques alternativos para estas prácticas, más allá de las pruebas funcionales.

Lo que lleva a concluir que el enfoque alternativo para la evaluación de la calidad de sistemas basados en ML se basa en la Integridad de sus Datos y la Robustez del Modelo propuestos por G. Fujii et al. [1].

Sobre la integridad de datos, A. Paka [21] y G. Fujii et al. [1] señalan que los datos deben estar limpios, ser apropiados, suficientes, consistentes y libres de imprecisiones a lo largo de su ciclo de vida. Sobre la robustez del modelo, S. Scher y A. Trügler [22], y G. Fujii et al. [1] señalan que los modelos deben ser adecuados, esto se refiere a que tan robusto es un modelo-

predicción de cara a perturbaciones específicas en los datos de entrada.

En cuanto a características de calidad, G. Fujii et al. [1], y N. Radziwill y M. Benton [23] señalan algunas de relevancia para la evaluación de este tipo de sistemas, comenzando por el desempeño, definido en términos de precisión y correctitud; la eficiencia, definida en términos de robustez y desempeño, y la explicabilidad e interpretabilidad, definidas como el grado con el que un ser humano puede comprender cómo la máquina toma decisiones con base en las salidas que provee y la efectividad definida en términos de funcionalidad. Reflexionando sobre métricas que permitan cuantificar algunas de estas características, para la robustez del modelo, la precisión puede medirse como el cociente del número de respuestas correctas entre el número de muestras; A. Paka [21] menciona tres tipos de problema en sistemas de ML en producción, datos ausentes, violación de rango, que se refiere a rangos numéricos no esperados, por ejemplo números negativos cuando se esperan números por encima del cero, y discrepancia en tipos de dato. Estos pueden constituir métricas de relevancia para la integridad de datos si se mide su porcentaje de ocurrencia.

La Figura 2 muestra los conceptos sobre SQA hasta ahora identificados y sus relaciones, siguiendo la notación de UML para Diagramas de clase.

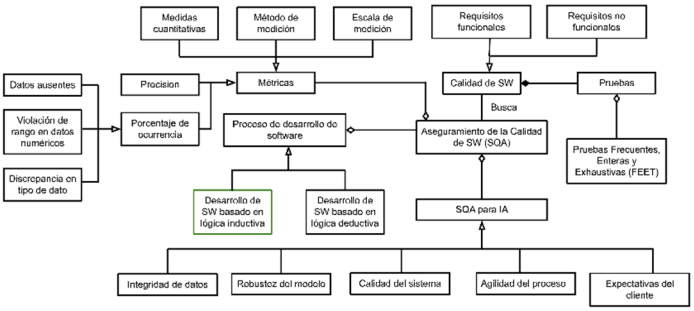

Figura 2: Modelo Conceptual de los Conceptos Expuestos sobre SQA. Fuente: Elaboración Propia

4. Aplicaciones Prácticas de la IA

Para profundizar sobre estrategias de prueba y SQA en ML fue necesario identificar los distintos tipos de aplicaciones de ML actualmente disponibles. IBM [24] y G. Fujii et al. [1] proponen una clasificación de las distintas aplicaciones de IA en tipos de aplicación, a continuación, se definen los más resaltantes.

Los sistemas generativos, que consisten en sistemas que generan contenido, es decir, imágenes, video, voz o texto, etc. mediante el empleo de técnicas como las redes contradictorias generativas (GANS), como afirma H. Saleem [25], en la cual se enfrentan dos redes neurales, una que genera contenido semejante a los datos de entrada y otra que diferencia entre los datos de entrada y los datos generados, con el objetivo de que la primera aprenda a generar data más real y la segunda aprenda a diferenciar mejor la data generada de la real. En la generación de textos se emplea en asistentes virtuales y *chatbots*, tanto en el campo del entretenimiento como en el de servicio al cliente.

Los sistemas de interfaz de usuario por voz (VUI), son sistemas que ejecutan instrucciones del usuario a partir del reconocimiento de voz, mediante el empleo de técnicas como Natural Language Processing (NLP) y Natural Language Understanding (NLU), y en los que se aplican tres procesos esenciales, reconocimiento de voz, que convierte la voz a texto, comprensión del lenguaje, que interpreta las instrucciones y sintetización de voz, que convierte la respuesta resultante en señales de voz. Su aplicación más relevante es en los altavoces inteligentes y asistentes virtuales como Siri y Alexa.

El procesamiento de imágenes, sistemas que emplean tecnologías de ML para el análisis de imágenes y video, en el campo de la conducción autónoma o en redes sociales, y en aplicaciones como AI-OCR (Object Character Recognition) que consiste en el reconocimiento de caracteres de texto a través de imágenes como, por ejemplo, en el llenado automático de formularios a través de fotografías de documentos.

Mientras que Fujii et al. [1] definen los sistemas generativos con un enfoque sobre la industria del entretenimiento, IBM [24] incluye parte de los sistemas generativos en los sistemas de reconocimiento de voz, en los cuales también se emplean técnicas como NLP y NLU y hacen referencia a los sistemas generativos en un enfoque de negocios al mencionar al servicio al cliente.

La Figura 3 muestra los conceptos sobre aplicaciones prácticas de la IA hasta ahora identificados y sus relaciones, siguiendo la notación de UML para Diagramas de clase.

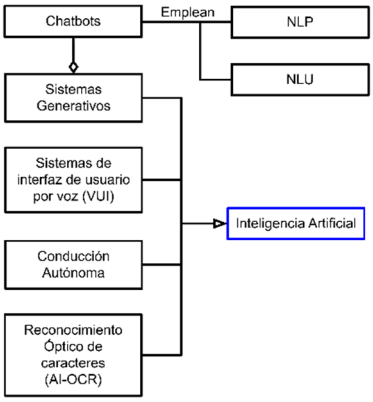

Figura 3: Modelo Conceptual de los Conceptos Expuestos sobre las Aplicaciones Prácticas de la IA. Fuente: Elaboración Propia

L. Fainchtein [26] y E. Kavlakoglu [27] definen NLP y NLU como técnicas que permiten a la máquina procesar el lenguaje humano, mediante el empleo de un conjunto de disciplinas como las estadísticas y técnicas de ciencia de datos; se diferencian principalmente en que en NLP el objetivo es procesar el texto o las palabras a partir de su formato de entrada, sea mediante texto o voz a un formato manejable por la computadora, mientras que el foco de NLU es la comprensión de este lenguaje con el fin de entender no solo el contenido, sino también la intención del usuario y procesar sus instrucciones de manera satisfactoria.

Oracle [28] y L. Fainchtein [26] definen los *chatbots* como aplicaciones que permiten al usuario entablar un diálogo con la máquina, con la idea de que esta interacción sea lo más humana posible, pudiendo estos ser sencillos y dar sólo respuestas a instrucciones básicas o habilitados por ML con la capacidad de refinar sus habilidades de conversación mediante el

entrenamiento. N. Radziwill y M. Benton [23] definen también la "humanidad" como un atributo de calidad relevante para esta aplicación de ML, definida como la capacidad de replicar el trato humano en la conversación.

Se puede, entonces, apreciar que las aplicaciones de ML abarcan un rango diverso de industrias y necesidades; en esta investigación el enfoque es sobre los sistemas generativos, más específicamente, aquellos relacionados con generación y/o procesamiento de textos (diálogos, preguntas, entre otros). La razón para definir este alcance, es que los sistemas generativos poseen un menor grado de complejidad, al ser el formato más básico de entrada y salida de datos en comparación a otros más complejos, como los sistemas de procesamiento de imágenes.

5. Modelo Conceptual Integrado

La Figura 4 muestra el mapa integrado de todos los conceptos hasta ahora identificados y sus relaciones, siguiendo la notación de UML para Diagramas de clase. Nótese que los conceptos vinculantes de los tres tópicos analizados son IA y el desarrollo de software basado en lógica inductiva.

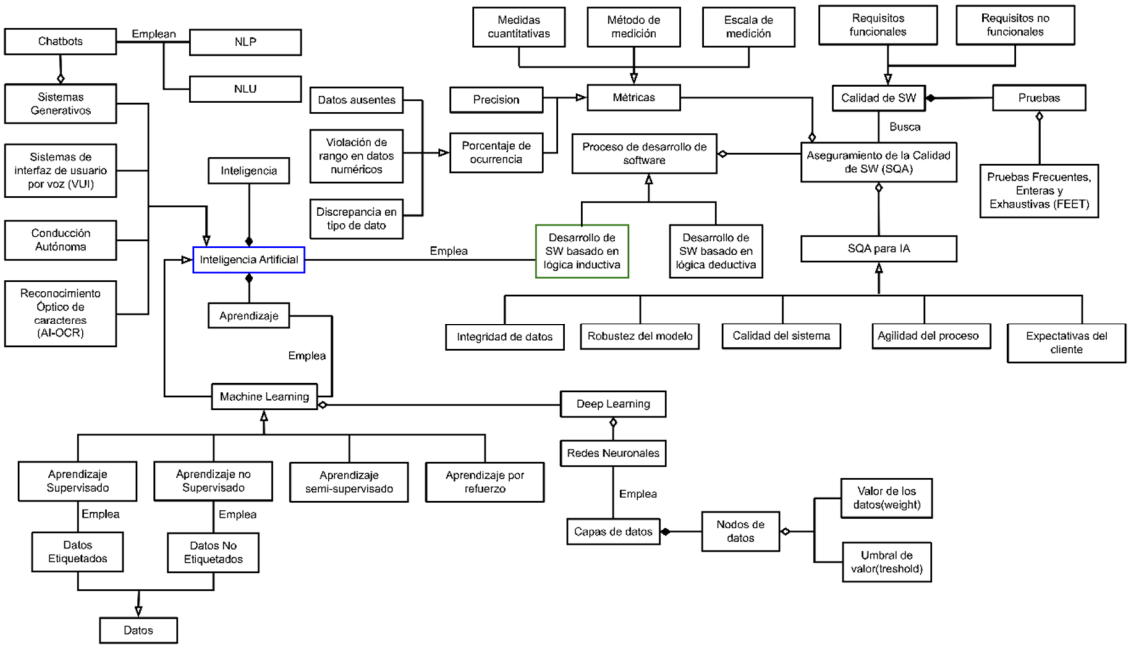

Figura 4: Modelo Conceptual Integrado de los Conceptos Expuestos en el Trabajo. Fuente: Elaboración Propia

6. Conclusiones

Este trabajo propone un modelo conceptual desarrollado siguiendo el Framework Metodológico Sistémico en su fase Diagnóstico [3], con el objetivo de apoyar la definición de una estrategia efectiva para la implementación de SQA en sistemas generativos impulsados por IA. El enfoque alternativo para aplicar SQA a software impulsado por ML se centrará en dos elementos como son la integridad de los datos de aprendizaje y de uso y la robustez del modelo.

7. Recomendaciones

Continuar con el desarrollo de la propuesta de una estrategia para SQA de sistemas generativos y su correspondiente evaluación. Así como también explorar a mayor profundidad posibles métricas que podrían ser adaptadas a este caso en particular con el objetivo de apoyar la evaluación de la calidad de estos sistemas.

Referencias

[1] G. Fujii, K. Hamada, F. Ishikawa, S. Masuda, S. Tokumoto, T. Toku, H. Ogawa, Y. Nishi, T. Myojin, M. Matsuya, K. Tsuchiya and Y. Ujita (Consorcio QA4A), *Guidelines for Quality Assurance of Machine Learning-Based Artificial Intelligence. International Journal of Software Engineering and Knowledge Engineering*, Japan, Febrero 2020.

[2] A. Carleton, M. Klein, J. Robert, E. Harper, R. Cunningham, D. De Niz, E. Desautels, J. Foreman, J. Goodenough, J. Herbsleb, C. Holland, I. Ozkaya, D. Schmidt and F. Shull. *Architecting the Future of Software Engineering: A National Agenda for Software Engineering Research & Development*. 2021.

[3] M. Pérez, A. Grimán and T. Rojas, *A Systemic Methodological Framework for IS Research*, AMCIS, pp. 4374–4383, 2004.

[4] R. Sternberg, *Human Intelligence*. Encyclopedia Britannica, Diciembre 2020.

[5] J. McCarthy, *What is Artificial Intelligence?*, Noviembre 2007, http://jmc.stanford.edu/articles/whatisai/whatisai.pdf.

[6] B. Copeland, *Artificial Intelligence.* Encyclopedia Britannica, Marzo 2018.

[7] L. Crow, A. Crow, *Educational Psychology, Van Nostrand Reinhold Inc*., U.S., 1963.

[8] D. Coon, *Introduction to Psychology. Exploration and Application*, 3rd edition, West Pub. Co., 1983.

[9] T. Mitchell, *Machine Learning*, McGraw Hill. 1997.

[10] A. Samuel, *Some Studies in Machine Learning Using the Game of Checkers.* IBM Journal of Research and Development Volume 3, Issue 3. Julio 1959.

[11] S. Russell and P. Norvig, *Inteligencia Artificial: Un Enfoque Moderno*. 2da edición. Pearson Education, S.A., Madrid. 2020.

[12] Y. Bengio, I. Goodfellow and A. Courville, *Deep Learning*. MIT Press. 2015.

[13] IBM. *Documentación de SPSS Modeler*, 2021, https://www.ibm.com/es-es/products/spss-modeler.

[14] J. Evans, *The Psychology of Deductive Reasoning*. Psychology Press. 2015.

[15] S-H. Nienhyus-Cheng and R. de Wolf, *Foundations of Inductive Logic Programming*, Springer-Verlag, 1997.

[16] Organización Internacional de Normalización, *Sistemas de Gestión de la Calidad — Fundamentos y Vocabulario* (ISO 9000:2015), 2015.

[17] J. Juran and J. De Feo, *Juran's Quality Handbook. The Complete Guide to Performance Excellence*, 6ta edición, McGraw Hill, 2010.

[18] Organización Internacional de Normalización, *Calidad de Software y Datos* (ISO 25010:2011), 2011.

[19] R. Pressman, *Software Engineering*, 5th edition, McGraw Hill, 2020.

[20] M. Estayno, G. Dapozo, L. Cuenca y C. Greiner, *Modelos y Métricas para Evaluar la Calidad de Software*, WICC, 2009.

[21] A. Paka, *Why Data Integrity is Key to ML Monitoring*, 2021, https://towardsdatascience.com/why-data-integrity-is-key-to-ml-monitoring-3843edd75cf5.

[22] S. Scher and A. Trügler, *Robustness of Machine Learning Models Beyond Adversarial Attacks*, 2017, https://doi.org/10.48550/arXiv.2204.10046.

[23] N.Radziwill and M.Benton, *Evaluating Quality of Chatbots and Intelligent Conversational Agents*, 2017, https://doi.org/10.48550/arXiv.1704.04579.

[24] IBM, *Machine Learning*, 2020, https://www.ibm.com/cloud/learn/machine-learning.

[25] H. Saleem, *What is Generative AI and How Much Power Does it Have?*, 2020, https://developer.ibm.com/blogs/what-is-generative-ai-and-how-much-power-does-it-have/.

[26] L. Fainchtein, *Generative vs Retrieval Based Chatbots: A Quick Guide*, 2020, https://blog.cloudboost.io/generative-vs-retrieval-based-chatbots-a-quick-guide-8d19edb1d645.

[27] E. Kavlakoglu, *NLP vs NLU vs NLG: The Differences Between Three Natural Processing Language Concepts*, 2020, https://www.ibm.com/blogs/watson/2020/11/nlp-vs-nlu-vs-nlg-the-differences-between-three-natural-language-processing-concepts/.

[28] Oracle, *What is a Chatbot*, https://www.oracle.com/chatbots/what-is-a-chatbot/.