Gestión del cambio para la Sostenibilidad Socioeconómica de las Naciones

Rodríguez, J., González, M., y Riaño, J. (2025). La relevancia de la recopilación de datos para el entrenamiento de un chatbot. En Chirinos, Y., Rojas, D y Barbera, R. (2025). (Eds.), *Gestión del cambio para la sostenibilidad socioeconómica de las Naciones*. Fondo Editorial Servando Garcés. DOI:

Capítulo XVII

LA RELEVANCIA DE LA RECOPILACIÓN DE DATOS PARA EL ENTRENAMIENTO DE UN CHATBOT

THE RELEVANCE OF DATA COLLECTION FOR TRAINING A CHATBOT

Juan Fernando Rodríguez Medina

Estudiante de ciencias de datos en la Fundación Universitaria Compensar. Correo electrónico: juanfernandorodriguez@ucompensar.edu.co . ORCID: https://orcid.org/0009-0006-1827-9312

Mireya Fideligna González

Estudiante de ciencia de datos en la Fundación Universitaria Compensar. Correo electrónico: mfidelignagonzalez@ucompensar.edu.co ORCID: https://orcid.org/0009-0007-9914-8687

Jeysson Alexander Riaño Pascagaza

Docente en la Fundación Universitaria Compensar. Correo electrónico: jarianop@ucompensar.edu.co. ORCID: https://orcid.org/0000-0002-4350-5469

Identificación de proyecto

La relevancia de la recopilación de datos para el entrenamiento de un chatbot.

Objetivo de la investigación

Analizar algoritmos de entrenamiento para chatbots aplicables en la Fundación Universitaria Compensar, orientados a mejorar la calidad académica, abordando aspectos legales y éticos relacionados con el uso de datos educativos.

Descripción del contexto

Este proyecto nace del propósito de innovar, y trabajar en la calidad y excelencia a la par de mantener la alta competitividad en un mundo en constante cambio. La investigación se centra en el desarrollo de un chatbot académico capaz de responder preguntas recurrentes de los estudiantes, especialmente en áreas donde suelen tener mayores dificultades, como la estadística y matemática en las áreas básicas. El propósito es apoyar el aprendizaje mediante el uso de herramientas de inteligencia artificial, haciendo que la interacción entre el estudiante y los contenidos académicos sean más entendibles y amigables en ese primer contacto con dichas asignaturas.

Para este propósito se analizaron diferentes algoritmos y herramientas que permiten entrenar un chatbot, incluyendo plataformas como RASA y bibliotecas como NLTK y TensorFlow. Estas herramientas permiten procesar lenguaje natural, lo que hace posible que el chatbot entienda y responda de manera coherente. Inicialmente se hará de forma independiente para pruebas piloto, y luego sea integrado a la plataforma educativa institucional.

Una parte fundamental del proyecto ha sido la recopilación de datos adecuados y confiables para el entrenamiento del chatbot. Para este proyecto se han investigado a profundidad a los aspectos legales y éticos que regulan el uso de datos en entornos educativos. En Colombia, actualmente hay varios proyectos de ley en curso, tal y como se observa en Colombia (2024), que buscan establecer marcos normativos para el uso de la inteligencia artificial. Por ello, el equipo ha considerado cuidadosamente estas regulaciones para asegurar que el chatbot cumpla con los principios de seguridad, privacidad y equidad.

Se reconoce que la investigación tiene la capacidad de generar cambios reales en la educación, en este caso, se busca que los resultados del proyecto puedan beneficiar directamente a la comunidad estudiantil, mejorando la calidad de la educación.

Revisión de la literatura

En entornos educativos la investigación sobre chatbot ha cobrado fuerza en los últimos años debido a la expansión del uso de la inteligencia artificial (IA) en múltiples sectores, entre estos, la educación superior. Los chatbots, definidos como programas capaces de simular una conversación humana, tienen su origen en la década de los años 60 con el desarrollo de ELIZA, considerado el primer chatbot de la historia hecho por Weizenbaum (1966). En un principio diseñados para responder a entradas mediante instrucciones establecidas, hoy en día se han transformado gracias al procesamiento de lenguaje natural (PLN) y el aprendizaje profundo, facilitando correspondencia más fluidas y flexibles.

McTear et al. (2016), definen los chatbots como sistemas que combinan IA y el Procesamiento de Lenguaje Natural (PLN), para facilitar la interacción con los usuarios a través de texto o voz. Por otra parte, Quiroga et al. (2020) señalan que, aunque su aplicación en áreas como el comercio y la atención al cliente es común, en el entorno formativo, aún existe un extenso potencial por investigar. Se destaca la relevancia de investigaciones como la presente, que buscan adaptar estas tecnologías al ámbito educativo. Variados instrumentos tecnológicos han promovido el desarrollo de chatbots. Entre ellas se destaca Natural Language

Toolkit (NLTK), una biblioteca de Python extensamente para tareas de PLN como tokenización, etiquetado gramatical y análisis semántico en NLTK Project (2023). Al igual, plataformas como RASA han adquirido notoriedad por su enfoque en chatbots personalizados y de código abierto, óptimos para implementaciones académicas.

La inteligencia artificial, en sus variantes de machine learning y deep learning, ha sido fundamental para avanzar en estas aplicaciones. Según Rodríguez et al. (2023). modelos como ChatGPT representan una nueva generación de asistentes virtuales, capaces de producir texto coherente y adaptado al contexto. Cervantes et al. (2024) refuerzan la anterior idea destacando el rol del aprendizaje profundo como base del desarrollo de modelos lingüísticos avanzados.

Desde una perspectiva ética y legal, enfatiza la importancia de considerar los derechos de autor y las licencias al utilizar contenido para entrenar modelos de IA, expuesto por Mendoza-Collazos (2025). En este sentido, la Ley de Inteligencia Artificial de la Unión Europea en (ArtificialIntelligenceAct.eu, s.f.) se establecen directrices claras sobre el uso, la transparencia y el riesgo asociado al despliegue de sistemas inteligentes. Estas normativas sirven de marco de referencia para iniciativas similares en América Latina y, particularmente, en Colombia, donde también se tramitan proyectos de ley relacionados con la IA teniendo en cuenta a Uribe (2024).

Metodología

La presente investigación teniendo en cuenta lo expuesto en el artículo de Bowen (2009) y parte del libro de Hernández et al. (2014) se adopta un enfoque cualitativo de tipo descriptivo con un enfoque en revisión documental. El propósito central es comprender y analizar cómo la recopilación, preparación y uso de datos inciden en el entrenamiento de chatbots académicos en la Fundación Universitaria Compensar. A través de este contexto se busca no solo describir el fenómeno, sino además interpretar sus implicaciones pedagógicas, tecnológicas, éticas y legales en el contexto colombiano, teniendo en cuenta antecedentes locales e internacionales, legislación vigente, implicaciones tecnológicas y de algoritmos.

La metodología se adaptó a las necesidades y se estructuró en tres fases principales:

La primera consistió en una revisión documental y teórica, en la cual se observaron fuentes académicas, artículos científicos, tesis de grado, normativas legales y documentos institucionales relacionados con inteligencia artificial, procesamiento de lenguaje natural (PLN) y chatbots educativos. Esta revisión permitió

identificar antecedentes, marcos normativos vigentes, proyectos que buscan crear una legislación entorno a la IA tanto de manera local, en Colombia, como a nivel internacional como se plasma en la Ley de IA en la Unión Europea (ArtificialIntelligenceAct.eu, s.f.).

Posteriormente, se realizó un análisis descriptivo con base en la información recopilada, se lleva a cabo un proceso de análisis orientado a describir las principales estrategias de entrenamiento de chatbots y las herramientas más utilizadas en este campo. Asimismo, se examinan los aspectos legales y éticos relacionados con la recopilación de datos, enfatizando en la protección de la privacidad, la seguridad de la información y la equidad en su uso.

Finalmente, se tiene la fase de aplicación contextual en la Fundación Universitaria Compensar, en la cual se reflexiona sobre la pertinencia de implementar chatbots académicos en áreas críticas como estadística y matemáticas básicas. Para ello, se describen los datos que deben recopilarse y prepararse y se plantean lineamientos técnicos que faciliten el diseño de un modelo conversacional accesible, escalable y ajustado a las necesidades de los estudiantes.

Hallazgos

Durante el desarrollo inicial del proyecto, se llevó a cabo un análisis de sondeo sobre las necesidades de nuestra universidad, Fundación Universitaria Compensar, respecto al impacto de la IA, donde se identificó la falta de herramientas inteligentes para la gestión, optimización, agilidad y competitividad del conocimiento. Respondiendo a esta necesidad, se han explorado diversos algoritmos de entrenamiento de chatbots, desde enfoques basados en reglas hasta modelos avanzados de aprendizaje automático y aprendizaje profundo (Deep Learning). También se ha evaluado la importancia de contar con bases de datos estructuradas y de calidad para garantizar un entrenamiento óptimo del modelo.

Desde el punto de vista técnico, se ha revisado el uso de bibliotecas y plataformas como Natural Language Toolkit (NLTK), Tensor Flow y RASA para el procesamiento de datos textuales. Asimismo, se han identificado formatos de contenido adecuados para el entrenamiento del chatbot, incluyendo texto, documentos estructurados, audios, imágenes y videos. Finalmente, se han considerado aspectos de seguridad y normativas como la Ley de Inteligencia Artificial de la Unión Europea y las licencias de software aplicables expuesto en (ArtificialIntelligenceAct.eu, s.f.). El 16 de diciembre de 2024, concluyó el primer periodo de la legislatura 2024-2025. En este momento en el Congreso se encuentran en trámite doce

proyectos de ley que buscan establecer un marco normativo sobre el uso de la Inteligencia Artificial (IA), los cuales vienen en curso desde la legislatura 2023-2024 expuesto en OlimpiA (2025). Cabe recordar que el ministro TIC, en el artículo de Jurídica (2024) anunció que radicaría el 30 de marzo de 2025, un proyecto de ley "sombrilla" que recogerá los principales puntos de los proyectos sobre inteligencia artificial.

Como parte de la continuidad del proyecto y facilitar el acceso a la información de los estudiantes de la Fundación Universitaria Compensar, y hacerlo de una manera más dinámica e interactiva, se ha definido que se enfocará específicamente en el desarrollo de chatbots aplicados a la asistencia académica, utilizando técnicas de procesamiento de Lenguaje Natural. Esta elección responde a la necesidad de mejorar la calidad y competitividad educativa. El chatbot estará diseñado para trabajar con lenguaje natural, aprovechando tecnologías como NLTK y RASA, que permiten interpretar y generar texto de forma automatizada. Desde su etapa inicial, el sistema tendrá la capacidad de responder tanto a preguntas específicas como generales, y para facilitar su funcionamiento, se incorporarán preguntas predeterminadas durante la fase piloto.

Se ha identificado que una de las materias con mayor índice de pérdida entre los estudiantes es Metamatemática fundamental, por lo que se reconoce la importancia de brindar apoyo académico en áreas de alta dificultad. Sin embargo, para facilitar la implementación el proyecto comenzará con un enfoque más acotado, centrado en el área de estadística, permitiendo entrenar el modelo con mayor precisión y controlar mejor su desempeño.

En cuanto a su estructura, en línea con lo propuesto por Díaz-González (2018), este chatbot tendrá una lógica conversacional estructurada y algoritmos de aprendizaje, similar a ChatGPT, lo que lo hace más intuitivo y natural para los usuarios. En la primera fase, funcionará de manera independiente como una herramienta de prueba, pero con el objetivo de integrarse posteriormente a la plataforma educativa de la institución. El sistema se iniciará como un modelo de pregunta y respuesta sin elementos interactivos avanzados como simulaciones o evaluaciones automatizadas.

Está pensado principalmente para estudiantes con un nivel de conocimiento principiante o intermedio en el área de matemáticas básicas, siguiendo el enfoque propuesto por Choque (2018) con el fin de reducir la carga del personal académico mediante un chatbot que pueda ofrecer orientación accesible y práctica extra, de las clases y tutorías que ofrece la universidad. El *framework opensource* elegido para su desarrollo e implementación será el de RASA Technologies GmbH (2025), por su enfoque específico en chatbots y su

facilidad de integración. Se desarrollará con una interfaz en línea, accesible desde cualquier navegador, y se trabajará con una API Restful que permita gestionar de manera eficiente la información. Todo esto usando el lenguaje de programación de Python y utilizando como algoritmo principal el DietClassifier y como alternativa el SpacyPipeline.

Como se observa en la Tabla 1, se consultó respecto a los tipos más comunes de chatbots, características principales, ejemplos y ventajas. Teniendo estas diferencias claras, se plantea qué tipo de chatbot es más viable y se adapta mejor al contexto académico en el que se quiere desplegar el mismo, se encuentra que el apropiado para las necesidades de la universidad, es el informativo.

Tabla 1

Cuadro comparativo de tipos de chatbots, así como sus características principales

Tipo de Chatbot	Propósito Principal	Ejemplos de Funciones	Ventajas
Informativo	Proporcionar información general.	Responder preguntas frecuentes, detalles de productos.	Reducción de consultas repetitivas, disponibilidad 24/7.
Transaccional	Facilitar transacciones y procesos.	Realizar compras, reservas, transferencias bancarias.	Eficiencia en procesos, conveniencia para el usuario.
Asistencial	Proporcionar soporte y asistencia personalizada.	Resolver problemas técnicos, guiar a través de procesos.	Asistencia adaptada, solución de problemas específicos.

Fuente: elaboración propia.

Esta tabla muestra las principales características de los tipos típicos de chatbots, ejemplos y principales ventajas. Finalmente, en etapas posteriores se evalúa la posibilidad de incorporar una interfaz gráfica que complemente la funcionalidad del sistema desplegándolo en la plataforma de la universidad.

Conclusiones

En conclusión, se puede afirmar que el desarrollo de un chatbot académico eficaz depende en gran medida de la calidad y estructura de los datos utilizados para su entrenamiento. Se concluye que contar con fuentes de información claras, organizadas y relevantes es esencial para garantizar que el chatbot ofrezca respuestas precisas, coherentes y útiles a los estudiantes. La recopilación y uso de datos en un proyecto de inteligencia artificial dentro de la institución educativa debe estar alineada con principios éticos y

normativas legales vigentes. En el contexto colombiano, se identificaron varios proyectos de ley en curso que abordan el uso responsable de la IA, lo que resalta la necesidad de que cualquier desarrollo tecnológico, como este chatbot, se realice con plena conciencia y respeto por la privacidad, la equidad y los derechos de los usuarios.

Se reconoce que la incorporación de tecnologías como los chatbots en la educación superior puede mejorar significativamente la experiencia de los estudiantes y facilitar el acceso al conocimiento. Sin embargo, para lograrlo, se requiere una estrategia integral que incluya fases de prueba, evaluación del impacto, retroalimentación constante y adaptación progresiva, comenzando por áreas específicas como la estadística y evolucionando hacia una integración más amplia en la plataforma educativa de la universidad.

Fuentes Bibliográficas

ArtificialIntelligenceAct.eu. (s.f.). L*ey de Inteligencia Artificial (AI Act)*. https://artificialintelligenceact.eu/es/

Bowen, G. A. (2009). Document Analysis as a Qualiative Research Method. *Qualitative Research Journal, 9*(2), 27-40. https://doi.org/10.3316/QRJ0902027

Cervantes Barrios, M., Montece Morante, G., y Mera Medina, E. (2024). Posibilidades de la inteligencia artificial en la educación. *Revista académica (Vol. 7, Núm. 4)*.

Choque Quispe, D. M. (2018). *Chatbot académico utilizando tecnologías cognitivas.* Lima. https://repositorioacademico.upc.edu.pe/bitstream/handle/10757/624418/CHOQUE_DM.pdf?sequence=5&isAllowed=y

Colombia, S. d. (2024). *Proyectos de Ley para uso responsable de IA y protección laboral ante nuevas tecnologías.* https://www.senado.gov.co/index.php/el-senado/noticias/5476-proyectos-de-ley-para-uso-responsable-de-ia-y-proteccion-laboral-ante-nuevas-tecnologias .

Díaz-González, D. (2018). *Diseño e implementación de un sistema de atención al cliente basado en un chatbot.* Trabajo de fin de grado / Tesis. https://repositorio.unican.es/xmlui/bitstream/handle/10902/16456/418113.pdf?sequence=1&isAllowed=y

Hernández Sampieri, R., Fernández Collado, C., y Baptista Lucio, P. (2014). *Metodología de la investigación* (6 ed.). McGraw-Hill.

Jurídica, P. (2024). *Así van los doce proyectos de Ley sobre inteligencia artificial (IA) en Colombia.* https://prensajuridica.com/details/item/32613-así-van-los-doce-proyectos-de-ley-sobre-inteligencia-artificial-ia-en-colombia.html

McTear, M., Callejas, Z., y Griol, D. (2016). *The conversational Interface.* Cham, Suiza: Springer International Publishing.

Mendoza-Collazos, J. (2025). *Periodico Unal*. https://www.emerald.com/qrj/article-abstract/9/2/27/360733/Document-Analysis-as-a-Qualitative-Research-Method?redirectedFrom=fulltext

NLTK Project. (2023). Natural Language Toolkit: https://www.nltk.org/

OlimpiA. (2025). *Ciberseguridad e identidad digital: Regulación de la inteligencia artificial en Colombia.* https://olimpiait.com/inteligencia-artificial/ciberseguridad-identidad-digital-regulacion-inteligencia-artificial-colombia/

Quiroga Pérez, J., Daradoumis, T., y Marquès Puig, J. M. (2020). Rediscovering the use of chatbots in education: A systematic literature review. *Computer Applications in Engineering Education, 28*(6), 1549-1565. https://doi.org/10.1002/cae.22326

Rasa Technologies GmbH. (2025). *Rasa Open Source Documentation*. https://legacy-docs-oss.rasa.com/docs/rasa/

Rodríguez Almazán, Y., Parra González, E., Zurita Aguilar, K., Mejía Miranda, J., y Bonilla Carranza, D. (2023). ChatGPT: la inteligencia artificial como herramienta de apoyo al desarrollo de las competencias STEM en los procesos de aprendizaje de los estudiantes. *Redalyc*. https://www.redalyc.org/journal/5122/512275598006/html/

Uribe Muñoz, A. (2024). *Por la cual se define y regula la inteligencia artificial, se ajusta a estándares de derechos humanos, se establecen límites frente a su desarrollo, implementación se modifica parcialmente la ley 1581 de 2012 y se dictan otras disposiciones.* https://www.camara.gov.co/inteligencia-artificial-2 .

Weizenbaum, J. (1966). ELIZA—A Computer Program for the Study of Natural Language Communication between Man and Machine. *Communications of the ACM, 9*(1), 36–45. https://doi.org/10.1145/365153.365168