

# Chatbot de Atención Estudiantil Basado en IA para la Universidad Jorge Tadeo Lozano

Sara Fiorella Ospina Pachón

sara.ospinap@utadeo.edu.co

Universidad Jorge Tadeo Lozano — Programa: Ingeniería de Sistemas

Curso: Inteligencia Artificial — Bogotá D.C., Colombia

5 de octubre de 2025

## Resumen

Este proyecto presenta el desarrollo de un chatbot inteligente para la atención estudiantil en la Universidad Jorge Tadeo Lozano. Su propósito es responder preguntas frecuentes sobre matrículas, horarios, asignaturas y servicios institucionales, optimizando la comunicación entre estudiantes y áreas administrativas. El sistema utiliza modelos de lenguaje basados en Transformers (BERT, T5-small) y técnicas de Procesamiento del Lenguaje Natural (PLN) para comprender consultas y generar respuestas coherentes. El modelo se entrenará con datos institucionales y corpus públicos en español, esperando alcanzar una precisión superior al 85% en clasificación de intenciones y reducir un 60% los tiempos de respuesta frente a la atención humana tradicional.

## Problema local y motivación

En la Universidad Jorge Tadeo Lozano, gran parte del tiempo del personal administrativo se destina a responder consultas repetitivas sobre procesos académicos, inscripciones, pagos y notas. Esto genera congestión en los canales de atención y disminuye la satisfacción estudiantil. En el contexto de Bogotá, donde la digitalización universitaria avanza rápidamente, se propone implementar un chatbot conversacional como solución innovadora para automatizar respuestas, mejorar la eficiencia operativa y promover la transformación digital institucional.

## Dataset

El modelo se entrenará con un conjunto de datos híbrido: (1) información institucional (preguntas frecuentes y guías académicas de la UJTL), y (2) datasets públicos “ChatterBot Corpus (ES)” y “FAQ dataset in Spanish”, con aproximadamente 18.000 instancias. El formato es CSV/JSON con columnas (‘intent’, ‘utterance’, ‘response’) y licencia MIT/Creative Commons. El dataset es válido por su diversidad, actualidad (2024–2025) y representatividad del contexto educativo colombiano.

## Tarea de IA y algoritmo(s)

La tarea corresponde a la clasificación de intenciones y generación de respuestas en texto (modalidad texto–texto). Se utilizarán los modelos BERT Multilingual Base para detección de intención y T5-small o DialoGPT para generación contextual. Estos modelos son apropiados por su capacidad de comprensión contextual y eficiencia con corpus limitados. Como línea base se empleará un modelo de regresión logística con vectores TF-IDF para comparación.

## Metodología y evaluación

El preprocesamiento incluirá limpieza, tokenización, eliminación de stopwords y lematización en español. Los datos se dividirán en 70% entrenamiento, 15% validación y 15% prueba. Se utilizarán métricas como Accuracy, F1-score, Precision/Recall y BLEU score, con validación cruzada (5-fold CV). La implementación se realizará con Python (Hugging Face, scikit-learn) y una interfaz web en Streamlit.

## Resultados esperados, ética y cronograma

Se espera que el chatbot alcance una F1-score  $\geq 0.85$ , responda correctamente al 90% de las preguntas frecuentes y reduzca los tiempos de atención en más del 50%. Desde la perspectiva ética, se garantizará la privacidad y anonimato de los usuarios, uso de datos con consentimiento informado y revisión humana para evitar sesgos o respuestas inadecuadas. Cronograma: Semana 1 - recolección de datos; Semana 2 - entrenamiento; Semana 3 - pruebas; Semana 4 - implementación y documentación.

## Roles del equipo y recursos

Sara Fiorella Ospina Pachón: liderazgo técnico, modelado y documentación. Integrante 2: adquisición y limpieza de datos. Integrante 3: desarrollo de interfaz en Streamlit. Recursos requeridos: GPU de entrenamiento (Google Colab), librerías Python (Transformers, Scikit-learn, Streamlit). Riesgos técnicos: limitación del dataset, sesgos lingüísticos y errores semánticos; mitigación mediante revisión manual y pruebas piloto controladas.

## Referencias

1. Devlin, J. et al. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL.
2. Wolf, T. et al. (2020). Transformers: State-of-the-Art Natural Language Processing. EMNLP.
3. Radford, A. et al. (2020). Language Models are Few-Shot Learners. OpenAI.
4. Google Cloud (2024). AI Conversational Platforms for Education.
5. Universidad Jorge Tadeo Lozano (2024). Portal Estudiantil y Guías Académicas.
6. Hugging Face (2025). Model Hub: BERT Multilingual Base.
7. López, M., & Gómez, P. (2023). Chatbots en educación superior colombiana. Rev. Innovación Digital.
8. Turing, A. (1950). Computing Machinery and Intelligence.
9. Pérez, A. (2022). Procesamiento de lenguaje natural con Python y SpaCy. Alfaomega.
10. MIT License (2024). ChatterBot Corpus (ES).