

# Metodología BRAIN: una metodología para la construcción de chatbots basados en BERT y sus aplicaciones en empresas

Miguel Angel González Reyes, Francisco Edgar Castillo Barrera,  
Héctor Gerardo Pérez González

Universidad Autónoma de San Luis Potosí,  
Facultad de Ingeniería, San Luis Potosí,  
México

a328333@alumnos.uaslp.mx,  
{hectorgerardo, ecastillo}@uaslp.mx

**Resumen.** Este artículo presenta una metodología para la construcción de chatbots basados en el modelo BERT, destacando su aplicabilidad en empresas. Se describe el proceso técnico de desarrollo y se analizan casos de uso en el ámbito empresarial, como la atención de solicitudes de empleo y la provisión de información sobre vacantes. Además, se propone un esquema metodológico que puede ser utilizado como referencia en la implementación de chatbots con inteligencia artificial.

**Palabras clave:** BERT, NLP, chatbot, inteligencia artificial, empresas, hugging face.

## BRAIN Methodology: A methodology for Building BERT-based Chatbots and their Applications in Business

**Abstract.** This article presents a methodology for building chatbots based on the BERT model, highlighting its applicability in business. The technical development process is described and business use cases are analyzed, such as handling job applications and providing information on vacancies. Furthermore, a methodological framework is proposed that can be used as a reference in the implementation of chatbots with artificial intelligence.

**Keywords:** BERT, NLP, chatbot, artificial intelligence, business, hugging face.

### 1. Introducción

En la última década, los avances en inteligencia artificial y procesamiento de lenguaje natural (PLN) han permitido el desarrollo de sistemas conversacionales cada vez más sofisticados. En particular, los modelos basados en arquitecturas transformer

[7], como BERT (Bidirectional Encoder Representations from Transformers), han mostrado resultados sobresalientes en tareas de comprensión y generación de lenguaje.

Este artículo presenta BRAIN (Base Extraction, Refinement, Answering, Information Retrieval, Natural Interaction), una metodología paso a paso para la construcción de chatbots basada en BERT, diseñada específicamente para aplicaciones en entornos empresariales. La metodología busca integrar herramientas como Hugging Face [6], SpaCy [8] y Sentence Transformers [7] para permitir una implementación eficiente, precisa y adaptable de chatbots que puedan interactuar con documentos institucionales, como vacantes o manuales internos.

A lo largo del trabajo se detallan las cinco fases de la metodología BRAIN, su implementación en Python, los resultados obtenidos en un caso de uso real, y un análisis comparativo respecto a metodologías existentes. Además, se presentan métricas de evaluación y una discusión sobre el impacto y las oportunidades de mejora del enfoque propuesto.

## **2. Metodología BRAIN**

La metodología propuesta se compone de cinco fases secuenciales que permiten transformar información documental en respuestas conversacionales contextualizadas. A continuación, se describe en detalle cada componente de la metodología BRAIN:

### **2.1 Base data extraction**

Esta fase se enfoca en la recopilación y extracción del conocimiento fuente desde archivos institucionales. En el caso de uso, se procesaron documentos PDF con descripciones de puestos y requisitos laborales mediante la biblioteca PyPDF2. El texto se segmentó por párrafos y se descartaron líneas vacías o irrelevantes. Esta información constituye el corpus base sobre el cual se construye el modelo conversacional.

### **2.2 Refinement of text**

Una vez extraída la información, se procede a su limpieza y normalización. Se utilizaron técnicas de normalización como la eliminación de acentos y conversión a minúsculas. Luego, se aplicó el modelo preentrenado all-MiniLM-L6-v2 de Sentence Transformers [7] para convertir cada párrafo en un vector semántico. Esto permite evaluar similitudes conceptuales más allá de la coincidencia literal de palabras.

### **2.3 Answer model with BERT**

Se utilizó el modelo de Hugging Face [6], para implementar una arquitectura de pregunta-respuesta en español. Aunque en la versión actual del sistema el modelo no se invoca directamente para generar respuestas, su incorporación está planeada como una mejora futura para afinar la selección de fragmentos contextuales.

## **2.4 Information retrieval strategy**

El sistema implementa una estrategia híbrida para recuperar la información más relevante:

**Coincidencia exacta.** Si una pregunta coincide textualmente con algún fragmento del corpus, se recupera esa sección.

**Heurística por intersección de palabras clave.** si no hay coincidencia exacta, se extraen palabras clave de la pregunta y se comparan con cada párrafo del corpus. El párrafo con mayor número de palabras coincidentes se selecciona como candidato.

Esta técnica permite responder incluso a preguntas mal formuladas o incompletas, maximizando la utilidad del chatbot en situaciones reales.

## **2.5 Natural interaction implementation**

Finalmente, se desarrolló una interfaz interactiva en Python donde los usuarios pueden formular preguntas en lenguaje natural. La interacción se realiza mediante consola, permitiendo que el sistema procese la entrada en tiempo real y devuelva una respuesta basada en los criterios anteriores. Se prevé extender esta funcionalidad mediante interfaces web o APIs REST para mejorar su usabilidad en entornos empresariales.

## **3. Resultados y evaluación**

Aunque la metodología BRAIN fue concebida originalmente para su uso en contextos empresariales (por ejemplo, atención a candidatos en procesos de reclutamiento), su primera implementación funcional se utilizó en un dominio académico. El objetivo fue validar su rendimiento empleando como dominio temático la orientación vocacional para aspirantes a los programas educativos que se ofrecen en la Universidad Autónoma de San Luis Potosí (UASLP).

### **3.1 Contexto de la evaluación**

Se utilizó como fuente de conocimiento un documento PDF fabricado a partir de documentación oficial que describe dos programas educativos: Ingeniería en Computación e Ingeniería en Sistemas Inteligentes. El texto incluye preguntas frecuentes sobre áreas de énfasis, perfil de ingreso y egreso, oportunidades laborales, movilidad y requisitos de admisión.

### **3.2 Metodología de evaluación**

Se realizaron 15 pruebas consecutivas, simulando preguntas frecuentes que podrían formular aspirantes reales sin conocimientos técnicos, incorporando intencionalmente lenguaje coloquial característico de una conversación casual y errores ortográficos, con el fin de verificar el funcionamiento del sistema.



**Fig. 1.** Metodología BRAIN, representada como una secuencia de pasos para la creación de chatbots basados en el modelo BERT.

### 3.3 Criterio de respuesta

El sistema procesa la entrada buscando:

- Coincidencia exacta del texto en los párrafos normalizados.
- Si no encuentra coincidencia, emplea una estrategia de intersección de palabras clave.
- En ambos casos, devuelve el primer párrafo completo hasta el punto final (.) en el texto original.

### 3.4 Ejemplos de interacción simulada

#### 3.5 Observaciones

- El sistema fue capaz de entregar respuestas relevantes en 14 de las 15 pruebas realizadas.
- Las preguntas formuladas con errores ortográficos o sin signos fueron correctamente entendidas.
- Las respuestas son literales del texto fuente, lo que garantiza su precisión.
- El sistema es especialmente eficaz para consultas concretas, pero podría mejorarse en respuestas abiertas implementando un sistema de validación por refuerzo.

## 4. Discusión

La implementación de la metodología BRAIN en un caso académico permite reflexionar sobre su eficacia, versatilidad y posibles limitaciones. El objetivo original del proyecto fue diseñar un chatbot de asistencia académica, orientado a apoyar a aspirantes universitarios en su proceso de orientación vocacional. Por ello, las pruebas

**Table 1.** Ejemplos de búsquedas realizadas con alumnos de nuevo ingreso que les hubiera gustado poder consultar sus dudas cuando eran aspirantes universitarios en la UASLP

Pregunta del usuario.	Respuesta obtenida
que ase un ingeniero en sistemas	Un ingeniero aplica conocimientos científicos y tecnológicos para resolver problemas.
que es un sistema inteligente	Los sistemas inteligentes son aquellos que perciben y responden al mundo que los rodea utilizando dispositivos computacionales.
hay trabajo para esa carrera	¡Sí, bastante! Hay una demanda muy grande a nivel mundial de profesionales de la computación, y con la llegada de la Industria 4.0, se requieren cada vez más ingenieros capaces de crear sistemas inteligentes.
en donde se puede trabajar	¡En donde tú quieras! A diferencia de otras carreras, las computadoras están y se necesitan en todos lados.
que materias tiene esa carrera	La disciplina de la computación es cada vez más amplia y diversa, con un rango de retos y oportunidades, por lo que hemos definido cuatro áreas de énfasis (o “tracks”), en donde te puedes especializar en ciertos temas en los últimos dos años de tu carrera.
necesito saber programar antes de entrar	No es necesario que sepas programar, pero si nunca lo has hecho te recomendamos que investigues lo que es la programación e intentes realizar algún programa sencillo, para que te des una idea de lo que implica.
como puedo estudiar en esa carrera	Para entrar a esta carrera deberás cumplir con los siguientes requisitos: Bachillerato completo en Ciencias Físico-Matemáticas, Bachillerato Tecnológico, Bachillerato General o Único.
que perfil necesito para entrar	Para ingresar y tener éxito en esta carrera es deseable que cuentes con: Razonamiento lógico y matemático Capacidad de iniciativa y liderazgo Capacidad de abstracción y síntesis Capacidad de razonamiento, análisis y comprensión de lectura.
cuales son los puestos donde puedo trabajar	Como ingeniero en sistemas inteligentes serás capaz de diseñar y desarrollar nuevos sistemas computacionales, por lo tanto, algunos puestos en donde te podrías desempeñar son: programador, analista de sistemas, administrador de redes.
que diferencia hay con otra ingeniería	La diferencia entre estas dos carreras es que tienen materias diferentes y enfoques diferentes, sistemas se basa en programación y computación en hardware.
que necesito para llegar a ser un ingeniero experimentado	Debes de practicar mucho.

de validación se enfocaron exclusivamente en este dominio, utilizando documentación oficial de programas educativos como fuente de conocimiento.

Durante el desarrollo de BRAIN, se identificó que la estructura modular propuesta —basada en la extracción de información, normalización semántica y recuperación híbrida de respuestas— tiene el potencial de adaptarse también a otros contextos más allá del académico. En particular, se reconoció su aplicabilidad en entornos empresariales para tareas como *la atención de solicitudes de empleo, programas de capacitación interna o soporte al cliente*, lo cual abre oportunidades de expansión futura para la metodología.

Uno de los principales hallazgos de la evaluación fue que la estrategia híbrida de recuperación de información —que combina coincidencia exacta e intersección de palabras clave— resultó altamente eficaz para responder preguntas comunes, incluso cuando estas contenían errores ortográficos o formulaciones ambiguas. Esto sugiere

**Table 2.** Tabla comparativa sobre proyectos recientes relacionados a chatbots con BERT y sus diferencias con la metodología BRAIN.

Referencia	¿Usa Metodología Formal?	Técnica Principal	Diferencias frente a BRAIN
Gopisetty (2024) [1]	No	Chatbot basado en BERT para e-commerce	Implementa sistema pregunta-respuesta para productos, pero no estructura modular definida; requiere entrenamiento específico para cada dominio. BRAIN es más generalizable y modular.
Babu & Boddu (2024) [2]	Parcial	Chatbot médico usando BERT fine-tuned	Utiliza fine-tuning en dominio médico, lo cual requiere datasets anotados. BRAIN evita fine-tuning pesado usando embeddings semánticos y búsqueda híbrida.
Singh et al. (2024) [3]	No	BERT combinado con Aprendizaje por Refuerzo (RL)	Enfocado en optimizar interacciones vía RL, no describe una metodología paso a paso; mayor complejidad técnica. BRAIN es más sencillo y directo para implementación práctica.
Hafidz et al. (2024) [4]	No	Chatbot de información turística con BERT	Utiliza BERT como motor de respuestas, pero sin un flujo estructurado ni estrategias híbridas de recuperación de información. BRAIN define un proceso modular claro.
Zeniarja et al. (2025) [5]	Parcial	Chatbot de salud mental usando BERT + ELECTRA	Integra dos modelos para mejorar comprensión emocional; enfoque de alta complejidad técnica. BRAIN está orientado a facilitar despliegues rápidos en contextos documentales bien definidos.

que, en dominios bien estructurados, sistemas basados en búsqueda semántica pueden ofrecer resultados satisfactorios sin necesidad de recurrir a modelos generativos complejos.

El uso de herramientas accesibles como Hugging Face [6], SpaCy [8] y Sentence Transformers [7] facilitó el desarrollo del sistema sin requerir grandes volúmenes de datos ni entrenamientos específicos, lo que disminuye las barreras de entrada para instituciones educativas o pequeñas organizaciones que buscan soluciones de atención automatizada.

No obstante, se reconocen algunas áreas de mejora. El sistema muestra limitaciones en la respuesta a preguntas abiertas o de carácter subjetivo, debido a su dependencia de fragmentos literales extraídos del corpus fuente. Futuras extensiones del sistema podrían considerar la incorporación de modelos BERT en la fase de refinamiento de respuestas, así como el uso de técnicas de aprendizaje por refuerzo para incrementar la precisión y adaptabilidad a distintas intenciones del usuario.

Finalmente, aunque los resultados preliminares en el dominio académico fueron positivos, será necesario realizar evaluaciones más amplias, incluyendo pruebas con usuarios con diferentes intereses y en diferentes contextos, para validar la efectividad y escalabilidad de la metodología BRAIN en entornos diversos.

Para reforzar la originalidad y aporte de la metodología BRAIN, se presenta una comparación con trabajos recientes que implementan chatbots basados en modelos BERT, pero que no necesariamente estructuran una metodología modular clara. La Tabla 2 muestra las principales diferencias:

Como puede observarse en la tabla 2, BRAIN muestra ventajas por su diseño modular, su accesibilidad para implementación sin entrenamiento intensivo y su estrategia híbrida de recuperación de información, lo que la hace especialmente

adecuada para contextos donde se dispone de corpus documentales definidos, pero no de grandes datasets anotados.

## 5. Conclusiones

La metodología BRAIN representa una propuesta estructurada, accesible y efectiva para el desarrollo de chatbots basados en modelos de lenguaje preentrenados como BERT. Su enfoque modular facilita la comprensión, implementación y adaptación en distintos dominios, ya sea empresarial, académico o institucional.

A través del caso de estudio presentado —un chatbot orientado a resolver dudas frecuentes de aspirantes universitarios sobre carreras en computación— se demostró que BRAIN permite construir sistemas funcionales, capaces de entregar respuestas relevantes sin necesidad de grandes volúmenes de datos ni entrenamiento supervisado. El sistema logró una tasa de éxito alta al recuperar información útil incluso en condiciones de formulación imperfecta por parte del usuario.

Se concluye que:

- El uso de modelos de embeddings [9] y estrategias de recuperación híbridas [10] puede ser suficiente para tareas de respuesta automática cuando el contexto está bien definido.
- La arquitectura propuesta permite extender la funcionalidad fácilmente, ya sea con nuevas fuentes de conocimiento o con modelos más sofisticados para el refinamiento de respuestas.
- Su diseño la convierte en una herramienta formativa útil, tanto para enseñar fundamentos de procesamiento de lenguaje natural como para desarrollar prototipos de soluciones conversacionales en el mundo real.

Como trabajo futuro, se buscará añadir al sistema que maneje BERT directamente, junto al sistema híbrido que ya se tiene para buscar obtener mejores resultados, se recomienda continuar con la validación de la metodología en escenarios más diversos, incluyendo interacciones con usuarios interesados en el contexto en que se quiera usar el ChatBot, así como explorar su integración con interfaces web y APIs para ampliar su aplicabilidad en entornos de producción.

**Agradecimientos.** Agradezco al Profesor Dr. Juan Carlos Cuevas Tello por sus enseñanzas y por sentar las bases de conocimiento que fueron fundamentales para el desarrollo de esta investigación.

## Referencias

1. Gopisetty, G., *Et al.*: Chatbot Building with BERT for E-Commerce. International Journal for Research in Applied Science and Engineering Technology, (2024) doi: 10.22214/ijraset.2024.59449.

2. Babu, A., Boddu, S.: BERT-Based Medical Chatbot: Enhancing Healthcare Communication Through Natural Language Understanding. *Exploratory Research in Clinical and Social Pharmacy*, 13. (2024) doi: 10.1016/j.rcsop.2024.100419.
3. Praneeth, KR., Ruprah, T.S., Madhuri, J.N.: Optimizing Customer Interactions: A BERT and Reinforcement Learning Hybrid Approach to Chatbot Development. *International Journal of Advanced Computer Science and Applications*, 15(9), (2024)
4. Hafidz, I., Mukti, B., Zahra, Q.: Chatbot Model Development Using BERT for West Sumatera Halal Tourism Information. *Halal Research Journal*, 4(2), (2024) doi: 10.12962/j22759970.v4i2.1819.
5. Zeniarja, J., Paramita, C., Subhiyakto, E.: Integrating ELECTRA and BERT Models in Transformer-Based Mental Healthcare Chatbot. *Indonesian Journal of Electrical Engineering and Computer Science*, 37(1), pp. 315–314 (2025) doi: 10.11591/ijeecs.v37.i1.pp315–324.
6. Osborne, C., Ding, J., Kirk, H.R.: The AI Community Building the Future? A Quantitative Analysis of Development Activity on Hugging Face Hub. *J Comput Soc Sc* 7, pp. 2067–2105 (2024) doi: 10.1007/s42001-024-00300-8.
7. Reimers, N., Gurevych, I.: Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, pp. 3982–3992 (2019) doi: 10.18653/v1/D19-1410.
8. Honnibal, M., Montani, I., Van Landeghem, S.: spaCy: Industrial-Strength Natural Language Processing in Python, (2020) doi: 10.5281/zenodo.1212303.
9. Mikolov, T., Chen, K., Corrado, G.: Efficient Estimation of Word Representations in Vector Space. *Proceeding of Workshop at ICLR*, (2013) doi: 10.48550/arXiv.1301.3781.
10. Karpukhin, V., Oguz, B., Min, S.: Dense Passage Retrieval for Open-Domain Question Answering. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pp. 6769–6781 (2020) doi: 10.18653/v1/2020.emnlp-main.550.