

# Análisis Visual de Problemas de Granularidad y Ambigüedad en el Agrupamiento de Intenciones con LLMs

**DATASET  
CLINC150**

Cecilia Vilca Alvites

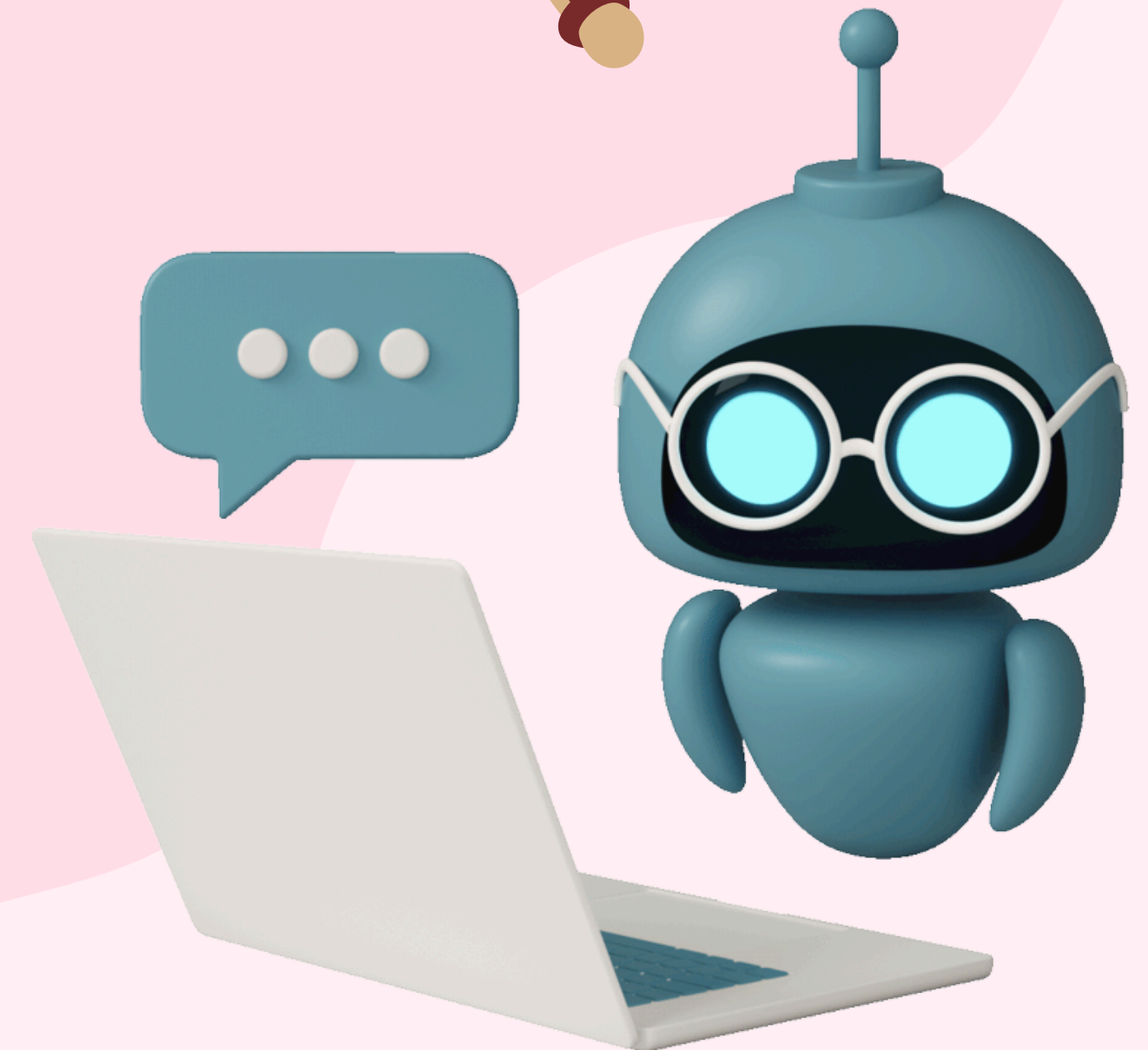


¿Qué es una Intención? El objetivo o propósito que un usuario tiene al interactuar con un sistema.

- Es la categoría que resume lo que el usuario quiere hacer

Ejemplo de intenciones:

- Enunciado del usuario: "Pon algo de rock clásico."
  - Intención: reproducir\_musica
- Enunciado del usuario: "¿Cuál es el pronóstico del tiempo para mañana en Arequipa?"
  - Intención: consultar\_clima



# Contexto

En la actualidad organizar grandes volúmenes de texto no etiquetado representa un desafío clave en el procesamiento de lenguaje natural. El agrupamiento de intenciones de usuario, común en asistentes virtuales, se ve afectado por la alta granularidad y ambigüedad semántica de las consultas. Aunque los LLMs han mejorado la representación contextual del texto, sus procesos de clustering siguen siendo poco interpretables. Esta falta de claridad, sumada a la escasez de herramientas visuales interactivas, dificulta el análisis y diagnóstico de errores. Ante ello, se justifica el desarrollo de enfoques visuales que permitan explorar y comprender mejor los agrupamientos generados.





# Problema

El agrupamiento de textos en lenguaje natural, específicamente las intenciones de usuario, representa un desafío significativo en el ámbito del procesamiento de lenguaje no supervisado. Este desafío se fundamenta en dos propiedades de los conjuntos de datos de intenciones:

## Alta Granularidad:

- Muchas intenciones son específicas y semánticamente cercanas.
- Representaciones por LLMs no siempre discriminan intenciones similares.
- Resultado: solapamiento de intenciones en el espacio de embeddings.

## Ambigüedad Semántica:

- Frases iguales o similares etiquetadas con intenciones distintas.
- El significado depende fuertemente del contexto.

Esta dificultad se acentúa en contextos no supervisados, donde la ausencia de etiquetas limita la detección de errores. La escasez de herramientas visuales e interactivas dificulta el análisis, haciendo fundamental contar con recursos que faciliten la identificación de intenciones mezcladas y frases ambiguas.





# Objetivos

- Usar LLMs para encontrar temas importantes en los textos y así mejorar la forma en que se representan antes de agruparlos.
- Identificar textos inusuales dentro de los grupos, usando tanto el análisis de intenciones como la visualización.
- Implementar métricas internas para evaluar la calidad de los clústeres, considerando coherencia semántica y separación.
- Diseñar e implementar una interfaz visual que permita explorar los grupos de textos creados, cambiar cuántos grupos queremos ver y mejorar los resultados de manera interactiva.





# Referencias

- [1] R. Peng, Y. Dong, G. Li, D. Tian, and G. Shan, “TextLens: Large language models-powered visual analytics enhancing text clustering,” *Journal of Intelligent & Fuzzy Systems*, DOI: 10.1007/s12650-025-01043-y, Feb. 2025.
- [2] A. Petukhova, J. Carvalho, and N. Fachada, “Text Clustering with Large Language Model Embeddings,” *International Journal of Cognitive Computing in Engineering*, vol. 6, pp. 100–108, Dec. 2025.
- [3] N. Arias, P. Singh, and A. B. Imbert, “Visual Analytics for Fine-grained Text Classification Models and Datasets,” *arXiv preprint arXiv:2405.02980*, 2024.
- [4] L. K. Miller and C. P. Alexander, “Human-interpretable clustering of short text using large language models,” *Royal Society Open Science*, vol. 12, no. 2, pp. 241088, 2025.
- [5] S. Hamada, “Processing of Semantic Ambiguity Based on Words Ontology,” *Journal of Computer Science*, vol. 16, no. 1, pp. 1–9, 2020.



# Muchas Gracias

Por ver esta presentación

