

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

**DATASET  
CLINC150**

Cecilia Vilca Alvites



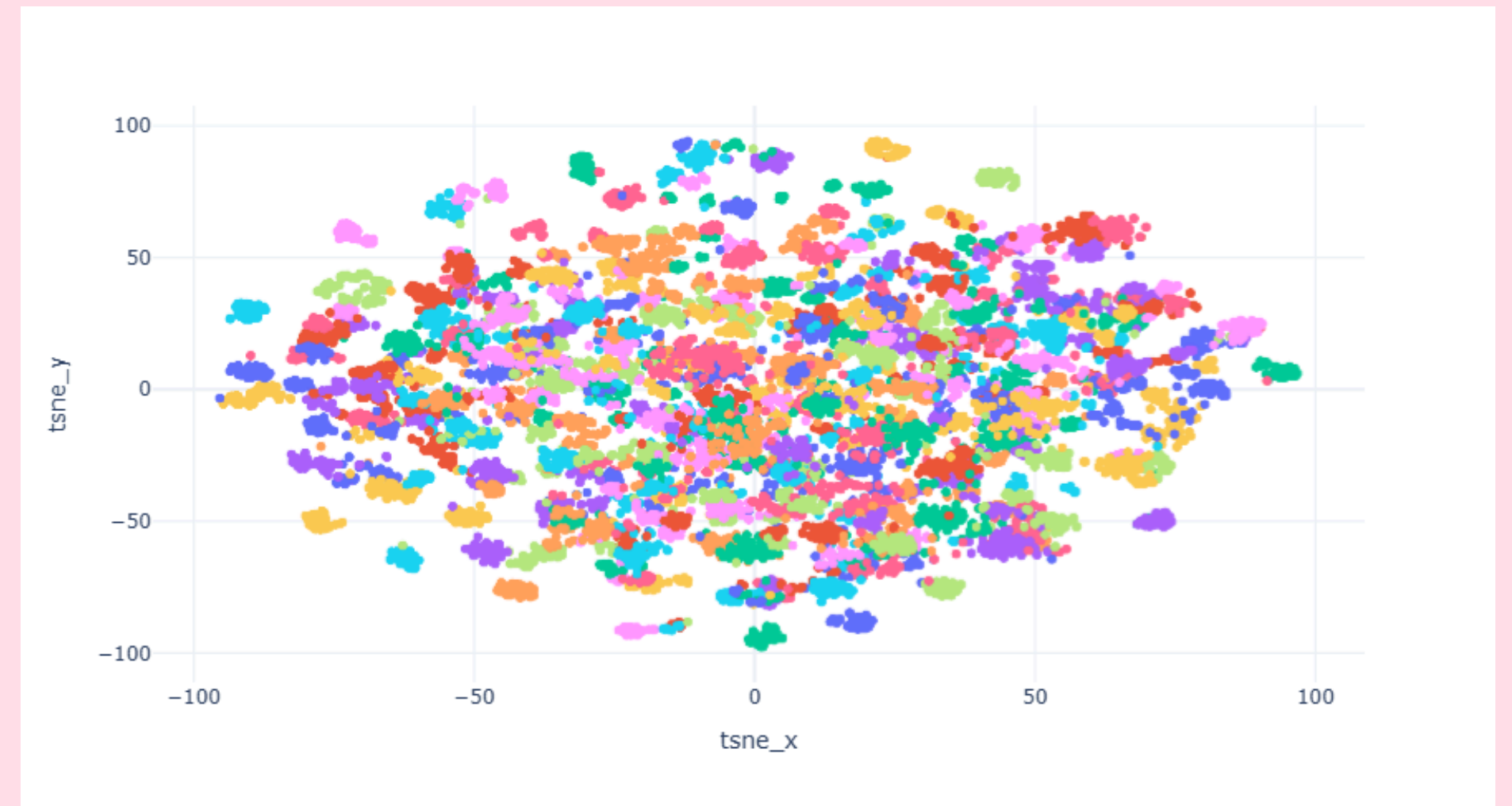
¿Qué es una Intención? El objetivo o propósito que un usuario tiene al interactuar con un sistema (ej. 'quiero pagar mi factura', 'dame el pronóstico del tiempo').

- Clasificar intenciones de usuario es fundamental para chatbots y asistentes virtuales.
- Pero, el lenguaje natural es ambiguo y multifacético.
- Problema 1: Granularidad: Demasiadas intenciones muy específicas.
- Problema 2: Ambigüedad: Consultas similares con intenciones diferentes.



# ¿Qué buscamos responder?

- Hipótesis 1 (Granularidad & Solapamiento): "La alta granularidad conduce a un solapamiento semántico significativo, manifestándose en patrones de agrupamiento complejos."
- Hipótesis 2 (Ambigüedad Semántica): "Existen instancias de ambigüedad semántica que provocan que frases idénticas o muy similares se asocien a intenciones distintas."
- Hipótesis 3 (Límites Difusos): "La ausencia de límites claros entre intenciones en el espacio de embeddings dificulta su separación efectiva."





# Metodología: Datos y Transformaciones

- Paso 1: Datos CLINC150: Dataset de consultas de usuario con 150 intenciones.
- Paso 2: Embeddings de Texto: Convertir texto a vectores numéricos (ej. con BERT, Sentence-BERT).
- Paso 3: Reducción de Dimensionalidad (t-SNE): De alta dimensión a 2D (tsne\_x, tsne\_y) para visualizar.
- Paso 4: Clustering K-Means: Agrupar los puntos en 'K' clusters.

TABLE I  
ATRIBUTOS PRINCIPALES DEL DATASET CLINC150 Y SU SIGNIFICADO

Atributo	Descripción	Tipo de Dato	Rango / Valores Posibles
text	Representa la <b>consulta original del usuario</b> en lenguaje natural. Es la entrada textual que el sistema debe procesar para inferir la intención.	String	Cadenas de texto que varían en longitud desde 2 hasta 136 caracteres. Ejemplos incluyen "what is my account balance" o "can you please tell me how much money I have left in my primary checking account after deducting all pending transactions?".
intent	Es la <b>etiqueta de la intención</b> subyacente asociada a la consulta de texto. Sirve como la verdad fundamental ( <i>ground truth</i> ) para la clasificación.	String	<b>150 categorías de intención</b> distintas y finamente granularizadas. Incluye intenciones como pay_bill, transfer, balance, greeting, goodbye, translate, money_transfer, y la categoría out_of_scope (OOS).
text_length	Atributo derivado que representa la <b>longitud de la consulta de texto</b> en número de caracteres. Se utiliza para análisis exploratorio.	Entero	Valores entre 2 y 136 caracteres. La mayoría de las consultas se agrupan alrededor de los 39 caracteres, con una mediana de 37.
split	Indica a qué subconjunto pertenece la instancia, utilizado para la división estándar del dataset para entrenamiento, validación y prueba de modelos.	String	train (entrenamiento), val (validación), test (prueba).
embeddings	Atributo derivado, no original del dataset, pero crucial para este proyecto. Son las <b>representaciones numéricas densas</b> de cada consulta de texto, generadas por un Large Language Model (LLM) (específicamente, all-MiniLM-L6-v2).	Vector	Vector de 384 dimensiones. Cada valor en el vector es un número flotante.
tsne_x, tsne_y	Atributos derivados de la reducción de dimensionalidad de los <i>embeddings</i> . Representan las <b>coordenadas 2D</b> de cada consulta en un espacio visual, obtenidas mediante t-SNE, facilitando su visualización e interpretación.	Flotante	Valores que varían según la distribución en el espacio 2D, generalmente en un rango de números reales.

# La Herramienta: Dashboard Interactivo



**Todo esto lo integramos en un Dashboard Interactivo hecho con Dash de Plotly. Esta herramienta nos permite no solo ver los resultados, sino también interactuar con ellos en tiempo real, lo que es vital para una exploración profunda.**

# Conclusiones

- Hipótesis 1 (Granularidad & Solapamiento): El scatter plot muestra mezclado de colores. La matriz de confusión revela que una intención se fragmenta en varios clusters (filas dispersas) y que un cluster agrupa varias intenciones (columnas mezcladas).
- Hipótesis 2 (Ambigüedad Semántica): El panel de detalles de selección, al revisar textos de consultas en clusters 'mixtos', evidencia frases muy similares para intenciones distintas.
- Hipótesis 3 (Límites Difusos): El scatter plot no muestra límites claros. Métricas como Silhouette Score (valores moderados) confirman que los clusters no están perfectamente separados. La matriz de confusión rara vez se vuelve diagonal pura, sin importar K.





# Muchas Gracias

Por ver esta presentación

