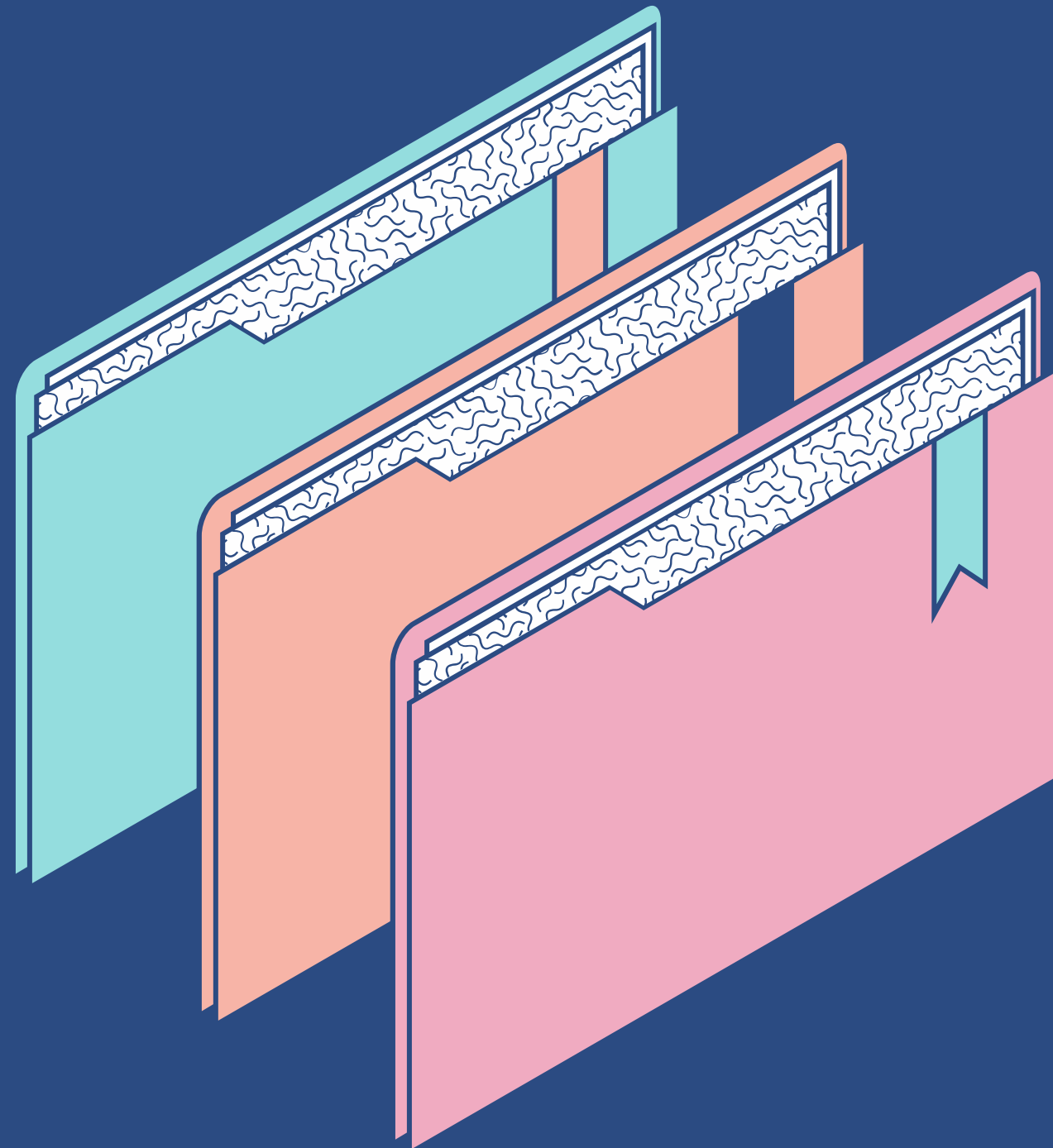




Implementación de RAG para la creación de estructuras cristalinas

Juan Sebastián Hernández G. Curso TAM 2024 II



Motivación

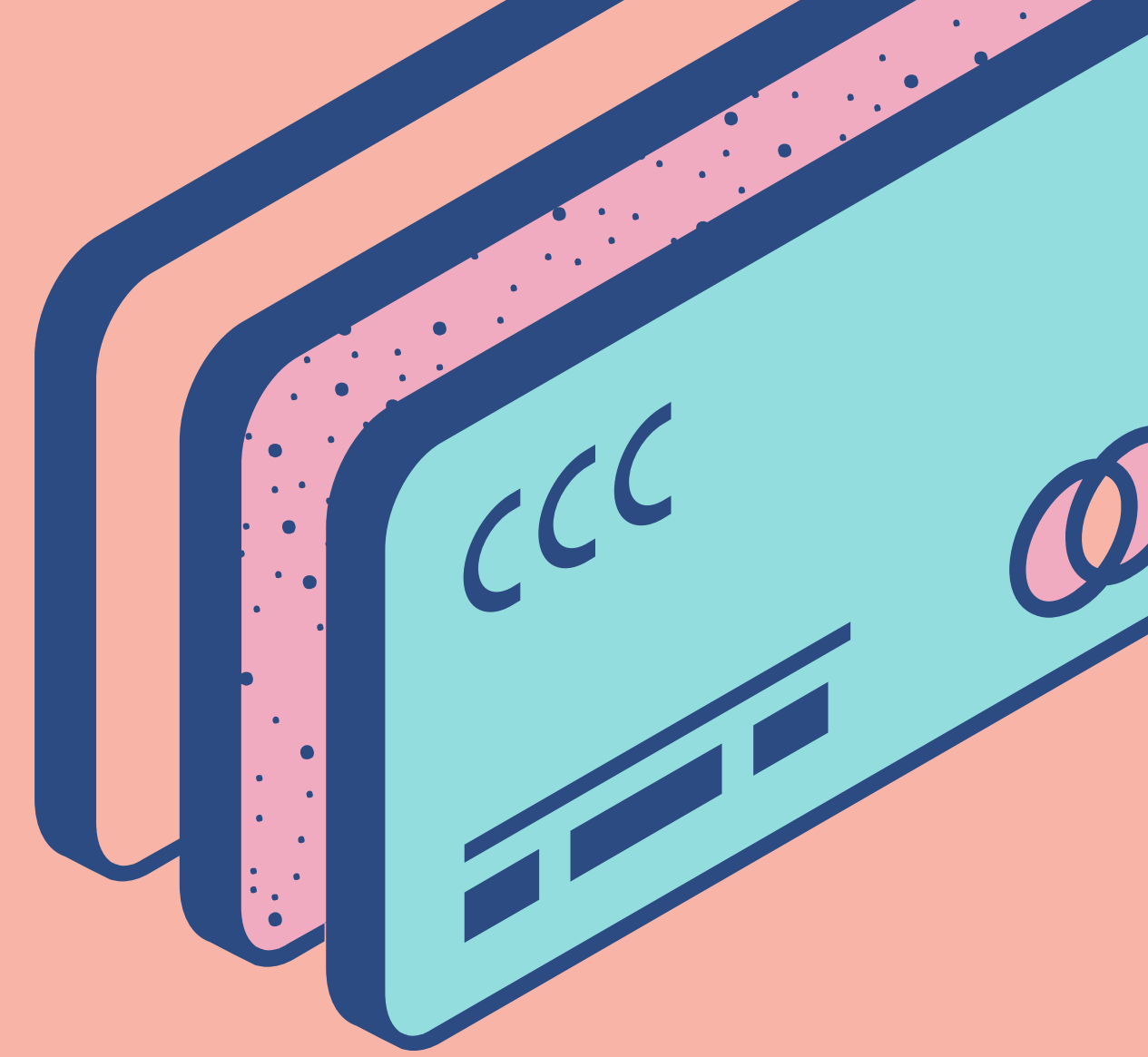
- La construcción de estructuras cristalinas es un paso clave en el desarrollo de simulaciones en diferentes disciplinas.
- Existen múltiples mecanismos para definir estructuras cristalinas que se pueden estudiar. [1]



Problema

- Las muestras dependen de la búsqueda en bases de datos que pueden ser limitadas o tediosas de recorrer.[1]
- Existen muchas propiedades y parametros que se deben conservar de forma precisa para asegurar que la muestra represente correctamente la realidad.[2]
- El formato de las bases de datos debe cambiarse para ajustarse a la necesidad de la simulación. [3]

Estado del arte



Implementación de una base de datos propia que recolecte muestras [3]

No elimina el problema de tener que revisar la base de datos

Uso de softwares especializados para cambiar el formato de los datos[4]

Limitado a la muestra exacta que se obtenga de la base de datos

Uso de modelos de IA para la predicción de materiales.[5]

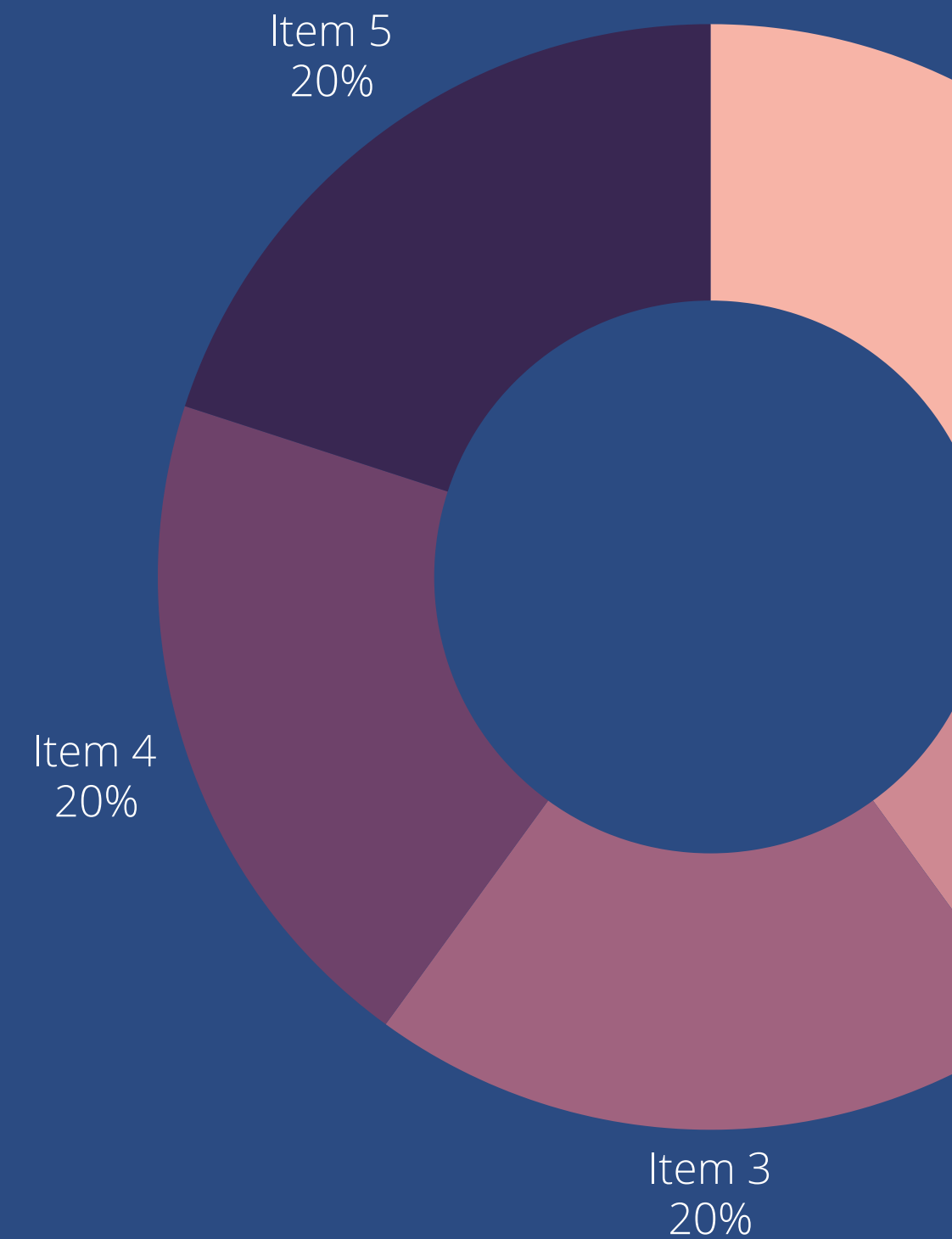
Principalmente usado para predecir propiedades pero no para definir las muestras.

Implementación

- Utilizando un modelo LLM ya existente, se automatizan las consultas a bases de datos conocidas, creando un sistema de generación aumentada por recuperación, el LLM obtiene contexto preciso de lo que se quiere construir.
- La información recuperada de la base de datos se pasa por un algoritmo de construcción de muestra ya definido que asegura el formato necesario para la aplicación para la que fue construido.
- El uso de librerías como langchain transforman al modelo de lenguaje en un agente capaz de buscar los materiales requeridos por el usuario y retornar una muestra construida acorde a los parámetros reales reportados

Resultados

- La implementación llevó a ver que los resultados dependen demasiado del modelo base.
- Si el agente no toma las acciones o no usa las herramientas que se le dan, es incapaz de responder a lo que se le pide.
- el acceso a un buen modelo fue el máximo limitante





Conclusiones

- En la teoría: la construcción del mecanismo RAG acelera mucho el proceso de búsqueda de parametros y asegura que el modelo tenga contexto y fundamento para la construcción de la muestra.
- Usar al modelo LLM como agente para ejecutar las búsquedas y el algoritmo libera carga y tiempo de trabajo.
- En la práctica: El metodo propuesto puede seguir presentando limitaciones cuando el material deseado presente cualidades muy especificas y dificiles de simular, o si carece de información suficiente reportada en las bases de datos
- La implementación es muy dependiente de la capacidad del LLM base para tomar acciones y usar las herramientas que se le proporcionan.

bibliografía

[1] D. Chen, W. Zhou, Y. Ji, C. Dong (2025). "Applications of Density Functional Theory to Corrosion and Corrosion Prevention of Metals: A Review." **Materials Genome Engineering**.

[2] Ward, Logan, et al. "Strategies for accelerating the adoption of Materials Informatics." *MRS Bulletin*, vol. 43, no. 9, Sept. 2018, pp. 683-689, <https://doi.org/10.1557/mrs.2018.204>.

[5] Liang, Jiechun, et al. "Accelerating perovskite materials discovery and correlated energy applications through artificial intelligence." *Energy Materials*, vol. 2, no. 3, Jan. 2022, p. 200016, <https://doi.org/10.20517/energymater.2022.14>.

[3] Ward, Logan, et al. "Strategies for accelerating the adoption of Materials Informatics." *MRS Bulletin*, vol. 43, no. 9, Sept. 2018, pp. 683-689, <https://doi.org/10.1557/mrs.2018.204>.

[4] Okadome Valencia, Hubert, et al. "New developments in the gdis simulation package: Integration of vasp and uspx." *Journal of Computational Chemistry*, vol. 42, no. 22, 7 June 2021, pp. 1602-1626, <https://doi.org/10.1002/jcc.26697>.