

Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias de la Computación

Doctorado en Ingeniería del Lenguaje y del Conocimiento



Propuesta de Proyecto de Tesis Doctoral

Modelo de Diálogo con IA Generativa Offline como Intermediario en la Automatización de Procesos en Laboratorios de Síntesis de Materiales.

Presenta

Alberto Esteban Reyes Peralta

Director de Tesis

Dr. David Eduardo Pinto Avendaño

Problemática

La automatización de laboratorios de investigación se ha convertido en una tendencia global en campos como la química, la biotecnología, la ingeniería de materiales y la medicina. Proyectos como los laboratorios autónomos de flujo dinámico en síntesis de materiales o las plataformas de alto rendimiento del National Renewable Energy Laboratory (NREL) demuestran que la combinación de robótica, instrumentación modular y algoritmos de aprendizaje automático permiten acelerar la experimentación, aumentar la calidad de los datos recolectados y reducir los costos de operación (NREL, 2025; Nishio et al., 2025). Sin embargo, persiste una brecha crítica: los sistemas de automatización carecen de una interfaz flexible e inteligente que permita a los investigadores comunicarse de manera natural con los equipos y traducir protocolos científicos en instrucciones reproducibles, seguras y validadas.

Este vacío se ha señalado en otros ámbitos científicos. Negrini y Lippi (2023), en su artículo “*Generative Artificial Intelligence in (Laboratory) Medicine: Friend or Foe?*”, destacan cómo la IA generativa podría asistir en tareas de laboratorio médico, como la interpretación de resultados o la redacción de informes. No obstante, advierten riesgos significativos asociados a errores de los modelos, falta de validación rigurosa, problemas éticos y de privacidad, así como una dependencia excesiva de herramientas aún inmaduras. Su análisis demuestra que, si bien la IA generativa es prometedora, no puede implementarse sin mecanismos de control robustos y sin un diseño especializado que garantice confiabilidad en contextos críticos.

En consecuencia, el problema central que aborda esta investigación es el diseño de un modelo de lenguaje grande (LLM) especializado que funcione como intermediario confiable entre investigadores y sistemas automatizados, permitiendo la ejecución de procesos experimentales en laboratorios de síntesis de materiales (y potencialmente en otros laboratorios) con mayor seguridad, trazabilidad y eficiencia.

Motivación y justificación

La motivación de este trabajo surge de la necesidad de dotar a los investigadores de herramientas que no solo aceleren la ejecución de experimentos, sino que reduzcan los errores derivados de la traducción manual de protocolos a lenguajes de control de equipos. En muchos laboratorios, la brecha entre el diseño experimental y la implementación técnica limita la productividad y retrasa descubrimientos que podrían tener impacto en áreas estratégicas como la energía, la nanotecnología o la biomedicina.

El uso de un LLM entrenado y adaptado al dominio de síntesis de materiales representa una oportunidad de frontera en la investigación científica, alineándose con tendencias internacionales en automatización y laboratorios autónomos (Aspuru-Guzik et al., 2018; Häse et al., 2019). A diferencia de sistemas de propósito general como ChatGPT, este modelo se diseñará para operar offline, lo que garantiza mayor seguridad, confidencialidad de datos sensibles y robustez en entornos experimentales.

La justificación de la propuesta se sostiene en dos ejes principales:

Científico-tecnológico: al explorar el desarrollo de un LLM especializado en automatización experimental, se contribuye a la frontera del conocimiento en inteligencia artificial aplicada, demostrando su viabilidad en un campo donde la precisión y la confiabilidad son críticas.

Socioeconómico: los avances derivados pueden replicarse en múltiples laboratorios (químicos, biomédicos, farmacéuticos) ampliando su impacto más allá de los materiales, beneficiando a sectores productivos y a la sociedad a través de descubrimientos más rápidos y reducción de costos experimentales.

Objetivo general

Diseñar e implementar un modelo de diálogo con IA generativa offline que funcione como intermediario en la automatización de procesos experimentales en laboratorios de síntesis de materiales, con el fin de mejorar la precisión, la eficiencia y la seguridad en la interacción entre investigadores y sistemas automatizados.

Objetivos específicos

- Analizar los principales retos y necesidades de interacción entre investigadores y sistemas automatizados en laboratorios de síntesis de materiales.
- Desarrollar un prototipo de modelo de lenguaje generativo offline capaz de interpretar instrucciones en lenguaje natural y traducirlas en comandos operativos para equipos automatizados.
- Evaluar el desempeño del modelo en términos de precisión, eficiencia y seguridad, comparándolo con métodos tradicionales de control e interacción.
- Validar la escalabilidad y adaptabilidad del modelo en otros tipos de laboratorios, explorando su potencial de generalización hacia diferentes áreas científicas.

Pregunta de investigación

¿Cómo puede la implementación de un modelo de diálogo con inteligencia artificial generativa offline asistir a los investigadores en su interacción con sistemas automatizados dentro de laboratorios de síntesis de materiales, mejorando la precisión experimental, la eficiencia operativa y la seguridad en comparación con los métodos tradicionales de comunicación humano-máquina?

Hipótesis

La implementación de un modelo de diálogo con inteligencia artificial generativa offline incrementa la precisión experimental, la eficiencia operativa y la seguridad en los procesos automatizados de los laboratorios de síntesis de materiales, en comparación con los métodos tradicionales de interacción humano-máquina.

Esta hipótesis se fundamenta en la premisa de que un modelo de lenguaje capaz de comprender y ejecutar instrucciones en lenguaje natural puede traducir las intenciones del investigador en acciones experimentales precisas, reduciendo errores humanos, optimizando el uso de recursos y minimizando la exposición a riesgos operativos. Asimismo, se espera que el modelo contribuya a la reproducibilidad de los resultados mediante la estandarización de los flujos de comunicación entre humanos y sistemas automatizados.

Referencias

- Aspuru-Guzik, A., Lindh, R., y Reiher, M. (2018). The Matter Simulation (R)evolution. *ACS Central Science*, 4, 144-152. <https://doi.org/10.1021/acscentsci.7b00550>
- Häse, F., Roch, L. M., y Aspuru-Guzik, A. (2019). Next-Generation Experimentation with Self-Driving Laboratories. *Trends in Chemistry*, 1, 282-291. <https://doi.org/10.1016/j.trechm.2019.02.007>
- Negrini, D., y Lippi, G. (2023). Generative Artificial Intelligence in (laboratory) medicine: friend or foe? *Biochimica Clinica*. https://doi.org/10.19186/BC_2023.025
- Nishio, K., Aiba, A., Takihara, K., Suzuki, Y., Nakayama, R., Kobayashi, S., . . . Hitosugi, T. (2025). A digital laboratory with a modular measurement system and standardized data format. *Digital Discovery*, 4, 1734-1742. <https://doi.org/10.1039/D4DD00326H>
- NREL. (2025). *Automated Data Analysis | Materials Science | NREL*. <https://www.nrel.gov/materials-science/automated-data-analysis>